

Large deviations of the trajectory of empirical distributions of Feller processes on locally compact spaces

Richard Kraaij ¹

May 29, 2015

Abstract

We study the large deviation behaviour of the trajectories of empirical distributions of independent copies of Feller processes on locally compact metric spaces. Under the condition that we can find a suitable core for the generator of the Feller process, we are able to define a notion of absolutely continuous trajectories of measures in terms of this core. Also, we define a Hamiltonian in terms of the linear generator and a Lagrangian as its Legendre transform.

The rate function of the large deviation principle can then be decomposed as a rate function for the initial time and an integral over the Lagrangian, finite only for absolutely continuous trajectories of measures.

The theorem partly extends the Dawson and Gärtner theorem [7], in the sense that it holds for diffusion processes where the drift and diffusion coefficients are sufficiently smooth. On the other hand, the result is sufficiently general to cover both Markov jump processes and discrete interacting particle systems [23].

1 Introduction

Dawson and Gärtner [7] proved the large deviation principle for the trajectory of empirical distributions of weakly interacting copies of diffusion processes. Additionally, they proved that the rate function can be decomposed as an entropy term for the large deviations at time zero and an integral over a quadratic Lagrangian, depending on position and speed. Recently, new proofs have been given in [2] using weak convergence methods and control theory and in [13] based on the comparison principle for an infinite dimensional Hamilton-Jacobi equation.

A similar results for Markov jump-processes, with a less explicit rate function, has been given in S. Feng [14], also see [15, 21]. Maes, Netočný and Wynants [3] study the large deviations of trajectories of the empirical distributions together with the empirical flow of a finite state space Markov jump process and give a Lagrangian form of the rate function.

¹Delft Institute of Applied Mathematics, Delft University of Technology, Mekelweg 4, 2628 CD Delft, The Netherlands.

These two sets of results raise the question whether a general approach is possible to prove large deviations for trajectories of weakly interacting, or even independent copies of processes with a ‘Lagrangian’ form of the rate function:

$$I(\gamma) := I_0(\gamma(0)) + \int_0^\infty \mathcal{L}(\gamma(s), \dot{\gamma}(s)) ds. \quad (1.1)$$

for absolutely continuous trajectories γ and infinity otherwise. Our main aim is to allow a large class of state spaces and processes including e.g. independent copies of whole interacting particle systems [23].

Feng and Kurtz [13] propose a general method to prove large deviations on the path-space based on the analogy of the large deviation principle to weak convergence. The approach involves the construction of a non-linear semigroup $\mathbf{V}(t)$ via the comparison principle for equations involving the generating Hamiltonian H and the Crandall-Liggett theorem [5]. Under some suitable conditions, they show that the semigroup $\mathbf{V}(t)$ can be re-expressed using a Nisio or Lax-Oleinik semigroup

$$\mathbf{V}(t)f(x) = \sup_{\gamma, \gamma(0)=x} f(\gamma(t)) - \int_0^\infty \mathcal{L}(\gamma(s), \dot{\gamma}(s)) ds$$

which in turn gives the Lagrangian form of the rate function.

In this paper, we give a proof of the large deviation principle for trajectories of empirical averages of independent copies of Feller processes on some space E without explicitly specifying the structure of the underlying process. Additionally, we express the rate function as in the form of (1.1).

We use the functional analytic structure underlying the large deviation principle introduced in [13], but our approach to the problem is different. The independence assumption implies that the large deviation principle can be proven via Sanov’s theorem and the contraction principle. Also, we can explicitly give the limiting non-linear semigroup $V(t)$ on E as $\log S(t)e^f$ where $S(t)$ is the semigroup of conditional expectations of the Markov process. This approach avoids the difficult problem of constructing a semigroup.

To obtain a Lagrangian form of the rate function, the main technical challenge is to show that $V(t)$ equals a Nisio semigroup $\mathbf{V}(t)$. The definition of the Nisio-semigroup poses us with two problems. First, we need a context-independent way to define absolutely continuous trajectories of measures, and secondly, we need a way to define a Lagrangian. To this end, we assume the existence of a suitable topology on a core of the generator $(A, \mathcal{D}(A))$ of the Feller process. The equality of $V(t)$ and $\mathbf{V}(t)$ is then proven using resolvent approximation arguments and Doob-transform techniques.

The rest of the paper is organised as follows. We start out in Section 2 with the preliminaries and state the two main theorems. Theorem 2.1 gives, under the condition that the processes solves the martingale problem, the large deviation principle. Under the condition that there exists a suitable core for the generator of the process, Theorem 2.8 gives the decomposition of the rate function.

In Section 3, we prove Theorem 2.1 using Sanov’s theorem for large deviations on the Skorokhod space and the contraction principle. We show that the rate function is given by a rate for the initial law, and a second term that is given as the supremum over sums of conditional large deviation rate functions. The

Legendre transforms of such conditional rate functions is given in terms of the non-linear semigroup $V(t)$. Additionally, we give a short introduction to the Doob transform, which we will use to study the non-linear semigroup.

In Sections 4 and 5, we prove Theorem 2.8. In the first section, we study the Hamiltonian, Lagrangian and a family of ‘controlled’ generators. Finally, in Section 5, we introduce the Nisio semigroup $\mathbf{V}(t)$ in terms of absolutely continuous trajectories and the Lagrangian, and show that it equals the non-linear semigroup $V(t)$.

In Section 6, we give three examples where Theorem 2.8 applies. We start with a Markov jump process. After that, we check the conditions for spatially extended interacting particle systems of the type that are found in Liggett [23]. Lastly, we check the conditions for a class of diffusion processes and show that, at least if the process is time-homogeneous and the diffusion and drift coefficients are sufficiently smooth, we recover the result for averages of independent and time-homogeneous processes by Dawson and Gärtner [7].

2 Preliminaries and main results

We start with some notation. Let (E, d) be a complete separable metric space with Borel σ -algebra \mathcal{E} . $\mathcal{M}(E)$ is the set of Borel measures of bounded total variation on E be equipped with the weak topology and $\mathcal{P}(E)$ is the subset of probability measures. We denote with $D_E(\mathbb{R}^+)$ the Skorokhod space of E valued càdlàg paths [12, Section 3.5], $\mathbb{R}^+ = [0, \infty)$. We write $\langle f, \mu \rangle$ for the integral of $f \in C_b(E)$ with respect to $\mu \in \mathcal{M}(E)$.

We define the relative entropy $H(\mu | \nu)$ of μ with respect to ν by

$$H(\mu | \nu) = \begin{cases} \int \log \frac{d\mu}{d\nu} d\mu & \text{if } \mu \ll \nu \\ \infty & \text{otherwise.} \end{cases}$$

On E , we have a time-homogeneous Markov process $\{X(t)\}_{t \geq 0}$ given by a path space measure \mathbb{P} on $D_E(\mathbb{R}^+)$. Let X^1, X^2, \dots be independent copies of X and let P the measure that governs these processes. We look at behaviour of the sequence $L_n := \left\{ L_n^{X(t)} \right\}_{t \geq 0}$,

$$L_n^{X(t)} := \frac{1}{n} \sum_{i=1}^n \delta_{\{X^i(t)\}},$$

under the law P . L_n takes values in $D_{\mathcal{P}(E)}(\mathbb{R}^+)$, the Skorokhod space of paths taking values in $\mathcal{P}(E)$. We also consider $C_{\mathcal{P}(E)}(\mathbb{R}^+)$ the space of continuous paths on $\mathcal{P}(E)$ with the topology inherited from $D_{\mathcal{P}(E)}(\mathbb{R}^+)$.

We say that L_n satisfies the large deviation principle (LDP) on $D_{\mathcal{P}(E)}(\mathbb{R}^+)$ with lower semi-continuous rate function $I : D_{\mathcal{P}(E)}(\mathbb{R}^+) \rightarrow \mathbb{R}^+$ if for every open set A

$$\liminf_{n \rightarrow \infty} \frac{1}{n} \log P[L_n \in A] \geq - \inf_{\mu \in A} I(\mu)$$

and for every closed set B

$$\limsup_{n \rightarrow \infty} \frac{1}{n} \log P[L_n \in B] \leq - \inf_{\mu \in B} I(\mu).$$

I is called good if its level sets $\{I \leq c\}$ are compact.

Suppose that $A : \mathcal{D}(A) \subseteq C_b(E) \rightarrow C_b(E)$, is a linear operator with a domain that separates points: for every $x, y \in E$, there exists a f in this set such that $f(x) \neq f(y)$. We say that X solves the martingale problem for $(A, \mathcal{D}(A))$ with starting measure \mathbb{P}_0 , if \mathbb{P}_0 is the law of $X(0)$ and if for every $f \in \mathcal{D}(A)$,

$$f(X(t)) - f(X(0)) - \int_0^t Af(X(s))ds$$

is a martingale for the natural filtration $\{\mathcal{F}_t\}_{t \geq 0}$ given by $\mathcal{F}_t = \sigma(X(s) \mid s \leq t)$. In Section 3, we obtain the following preliminary result.

Theorem 2.1. *Let X , represented by the measure \mathbb{P} on $D_E(\mathbb{R}^+)$ solve the martingale problem for $(A, \mathcal{D}(A))$ with starting measure \mathbb{P}_0 . Then, the sequence L_n satisfies the large deviation principle with good rate function I , which is given for $\nu = \{\nu(t)\}_{t \geq 0} \in D_{\mathcal{P}(E)}(\mathbb{R}^+)$ by*

$$I(\nu) = \begin{cases} H(\nu(0) \mid \mathbb{P}_0) + \sup_{\{t_i\}} \sum_{i=1}^k I_{t_i - t_{i-1}}(\nu(t_i) \mid \nu(t_{i-1})) & \text{if } \nu \in C_{\mathcal{P}(E)}(\mathbb{R}^+) \\ \infty & \text{otherwise,} \end{cases}$$

where $\{t_i\}$ is a finite sequence of times: $0 = t_0 < t_1 < \dots < t_k$. For $s \leq t$, we have

$$I_t(\nu_2 \mid \nu_1) = \sup_{f \in C_b(E)} \{ \langle f, \nu_2 \rangle - \langle V(t)f, \nu_1 \rangle \}, \quad (2.1)$$

where $V(t)f(x) = \log \mathbb{E} [e^{f(X(t))} \mid X(0) = x]$.

For further results, we introduce some additional notation. For a locally convex space (\mathcal{X}, τ) , we write \mathcal{X}' for its continuous dual space and $\mathcal{L}(\mathcal{X}, \tau)$ for the space of linear and continuous maps from (\mathcal{X}, τ) to itself. Also, for $x \in \mathcal{X}$ and $x' \in \mathcal{X}'$, we write $\langle x, x' \rangle := x'(x) \in \mathbb{R}$ for the natural pairing between x and x' . For two locally convex spaces \mathcal{X}, \mathcal{Y} and a continuous linear operator $T : \mathcal{X} \rightarrow \mathcal{Y}$, we write $T' : \mathcal{Y}' \rightarrow \mathcal{X}'$ for the adjoint of T , which is uniquely defined by $\langle x, T'(y') \rangle = \langle Tx, y' \rangle$, see for example Treves [27, Chapter 19]. For a neighbourhood \mathcal{N} of 0 in \mathcal{X} , we define the polar of $\mathcal{N}^\circ \subset \mathcal{X}'$ by

$$\mathcal{N}^\circ := \{u \in \mathcal{X}' \mid |\langle x, u \rangle| \leq 1 \text{ for every } x \in \mathcal{N}\}. \quad (2.2)$$

We say that a locally convex space \mathcal{X} is barrelled if every barrel is a neighbourhood of 0. A set S is a barrel if it is convex, balanced, absorbing and closed. S is balanced if we have the following: if $x \in S$ and $\alpha \in \mathbb{R}$, $|\alpha| \leq 1$ then $\alpha x \in S$. S is absorbing if for every $x \in \mathcal{X}$ there exists a $r \geq 0$ such that if $|\alpha| \geq r$ then $x \in \alpha S$. We give some background and basic results on barrelled spaces in Appendix 8.

To rewrite the rate function obtained in Theorem 2.1, we restrict to locally compact metric spaces (E, d) and we consider the situation where $S(t)f(x) = E[f(X(t)) \mid X(0) = x]$ is a strongly continuous semigroup on the space $(C_0(E), \|\cdot\|)$: for every $t \geq 0$, the map $S(t) : (C_0(E), \|\cdot\|) \rightarrow (C_0(E), \|\cdot\|)$ is continuous, and for every $f \in C_0(E)$, the trajectory $t \mapsto S(t)f$ is continuous in $(C_0(E), \|\cdot\|)$.

Let $(A, \mathcal{D}(A))$ be the generator of the semigroup $S(t)$. It is a well known result that X solves the martingale problem for $(A, \mathcal{D}(A))$ [12, Proposition 4.1.7], so the above result holds for the process $\{X(t)\}_{t \geq 0}$.

Our goal is to rewrite I as

$$I(\nu) = H(\nu(0) | \mathbb{P}_0) + \int_0^t \mathcal{L}(\nu(s), \dot{\nu}(s)) ds$$

for a trajectory ν of probability measures that is absolutely continuous in some sense. Thus our first problem is to define differentiation in a context for which no suitable structure on E or $\mathcal{P}(E)$ is known. Therefore, we will have to tailor the definition of differentiation to the process itself. Suppose that $\mu(t)$ is the law of $X(t)$ under \mathbb{P} . Then we know that $t \mapsto \mu(t) = S(t)' \mu(0)$ is a weakly continuous trajectory in $\mathcal{P}(E)$, so can ask whether for $f \in \mathcal{D}(A)$ the trajectory $t \mapsto \langle f, \mu(t) \rangle$ is differentiable as a function from $\mathbb{R}^+ \rightarrow \mathbb{R}$:

$$\frac{\partial}{\partial t} \langle f, \mu(t) \rangle = \frac{\partial}{\partial t} \langle S(t)f, \mu(0) \rangle = \langle S(t)Af, \mu(0) \rangle = \langle Af, \mu(t) \rangle. \quad (2.3)$$

So our candidate for $\dot{\mu}(t)$ would be $A'\mu(t)$, which is a problematic because $(A, \mathcal{D}(A))$ could be unbounded. To overcome this, and other problems, we introduce two sets of conditions on $(A, \mathcal{D}(A))$.

Recall that D is a core for $(A, \mathcal{D}(A))$ if for every $f \in \mathcal{D}(A)$, we can find a sequence $f_n \in D$ such that $f_n \rightarrow f$ and $Af_n \rightarrow Af$.

Condition 2.2. There exists a core $D \subseteq \mathcal{D}(A)$ dense in $(C(E), \|\cdot\|)$ that satisfies

- (a) D is an algebra, i.e. if $f, g \in D$ then $fg \in D$,
- (b) if $f \in D$ and $\phi : \mathbb{R} \rightarrow \mathbb{R}$ a smooth function on the closure of range of f , then $\phi \circ f - \phi(0) \in D$,

In the case that E is compact, $C_0(E) = C(E)$, then (b) can be replaced by

- (b') if $f \in D$ and $\phi : \mathbb{R} \rightarrow \mathbb{R}$ a smooth function on the range of f , then $\phi \circ f \in D$.

Note if a dense subspace $D \subseteq \mathcal{D}(A)$ satisfies $S(t)D \subseteq D$, then D is a core for $(A, \mathcal{D}(A))$ [12, Proposition 1.3.3].

Under Condition 2.2, we define the operator $H : D \rightarrow C_0(E)$ and for every $g \in D$ the operator $A^g : D \rightarrow C_0(E)$ by

$$\begin{aligned} Hf &= e^{-f} A e^f, \\ A^g f &= e^{-g} A (f e^g) - (e^{-g} A e^g) f. \end{aligned}$$

If E is non-compact, these definitions needs some care as $e^f \notin C_0(E)$. This can be solved by looking at the one-point compactification of E , see Section 4.1. In Section 4, we will show that $\{V(t)\}_{t \geq 0}$ turns out to be a non-linear semigroup on $C_0(E)$ which has H as its generator. The operators A^g are generators of Markov processes with law \mathbb{Q}^g on $D_E([0, t])$ that are obtained from \mathbb{P} by

$$\frac{d\mathbb{Q}_t^g}{d\mathbb{P}_t}(X) = \exp \left\{ g(X(t)) - g(X(0)) - \int_0^t Hg(X(s)) ds \right\}, \quad (2.4)$$

where \mathbb{P}_t and \mathbb{Q}_t^g are the measures \mathbb{P} and \mathbb{Q}^g restricted to times up to t , see Theorem 4.2 in [25].

Condition 2.3 (Conditions on the core). D satisfies Condition 2.2 and there exists a topology τ_D on D such that

- (a) (D, τ_D) is a separable barrelled locally convex Hausdorff space.
- (b) The topology τ_D is finer than the sup norm topology restricted to D .
- (c) The maps $\exp -1 : (D, \tau_D) \rightarrow (D, \tau_D)$ and $\times : (D, \tau_D) \times (D, \tau_D) \rightarrow (D, \tau_D)$, defined by $f \mapsto e^f - 1$, respectively $(f, g) \mapsto fg$ are continuous.
- (d) We have $S(t)D \subseteq D$, $V(t)D \subseteq D$ and the semigroups $\{S(t)\}_{t \geq 0}$ and $\{V(t)\}_{t \geq 0}$ restricted to D are strongly continuous for (D, τ_D) .
- (e) The map $A : (D, \tau_D) \rightarrow (C_0(E), \|\cdot\|)$ is continuous.
- (f) There exists a barrel $\mathcal{N} \subseteq D$ such that

$$\sup_{f \in \mathcal{N}} \|Hf\| \leq 1,$$

and for every $c > 0$

$$\sup_{f \in c\mathcal{N}} \|Hf\| < \infty.$$

Conditions (a) to (e) are related to the differentiation of the trajectories of measures that will turn up in our large deviation problem. Condition (a) implies that (D, τ_D) is well behaved as a locally convex space and, among other things, makes sure that we are able to define the Gelfand integral, see Appendix 8. Condition (b) implies that $(\mathcal{M}(E), wk)$ is continuously embedded in (D', wk^*) , so that every weakly continuous trajectory of measures can in fact be seen as a weak* continuous trajectory in D' . Important in this light is that the conditions in (d) on $\{S(t)\}_{t \geq 0}$ imply that $t \mapsto S(t)'\mu$ is weak* continuous in D' for all measures $\mu \in \mathcal{P}(E)$. (e) implies that we take the adjoint of $A : (D, \tau_D) \rightarrow (C_0(E), \|\cdot\|)$, so that we obtain a weak to weak* continuous map $A' : \mathcal{M}(E) \rightarrow D'$. Returning to Equation (2.3), we now have a good definition for $\dot{\mu}(t) := A'\mu(t) \in D'$. Furthermore, we can also differentiate trajectories of measures that are obtained from X via a tilting procedure, e.g. Equation (2.4) by Lemma 2.5 below. Condition (f) is the main technical condition on H which implies for example the compactness of the level sets of \mathcal{L} . Using these compactness arguments, we are able to rewrite I .

Remark 2.4. The conditions on $\{V(t)\}_{t \geq 0}$ in (d) seem to be too strong. They are used to prove that $V(t)$ equals $\mathbf{V}(t)$, see Proposition 5.10. If, for example, the range condition for H , i.e. $\overline{\text{Ran}(\mathbb{1} - \lambda H)(D)} = C_0(E)$ for $\lambda > 0$, can be checked directly, or if the comparison principle for H can be proven, the result of Proposition 5.10 would follow without $V(t)D \subseteq D$ or the strong continuity of $V(t)$.

The following lemma is a consequence of Condition 2.3 (c) and (e) and the proof is elementary.

Lemma 2.5. *Let (D, τ_D) satisfy Condition 2.3, then the maps $\mathcal{A} : (D, \tau_D) \times (D, \tau_D) \rightarrow (C_0(E), \|\cdot\|)$ given by $\Phi(g, f) = A^g f$ and the operator $H : (D, \tau_D) \rightarrow (C_0(E), \|\cdot\|)$ are continuous.*

Remark 2.6. The results of this paper also hold in the case that Condition 2.3 (c) fails as long as the conclusions of Lemma 2.5 hold. In all examples that we consider in Section 6 (c) is satisfied.

For the next definition we will need the Gelfand or weak* integral, which is introduced in Appendix 8, but the rigorous construction of this integral can be skipped on the first reading.

Definition 2.7. Define $D - \mathcal{AC}$, or if there is no chance of confusion, \mathcal{AC} , the space of absolutely continuous paths in $C_{\mathcal{P}(E)}(\mathbb{R}^+)$. A path $\nu \in C_{\mathcal{P}(E)}(\mathbb{R}^+)$ is called absolutely continuous if there exists a (D', wk^*) measurable curve $s \mapsto u(s)$ in D' with the following properties:

- (i) for every $f \in D$ and $t \geq 0$ $\int_0^t |\langle f, u(s) \rangle| ds < \infty$,
- (ii) for every $t \geq 0$, $\nu(t) - \nu(0) = \int_0^t u(s) ds$ as a D' Gelfand integral.

We denote $\dot{\nu}(s) := u(s)$. Furthermore, we will denote \mathcal{AC}_μ for the space of absolutely continuous trajectories starting at μ_0 , and \mathcal{AC}^T for trajectories that are only considered up to time T . Similarly, we define \mathcal{AC}_μ^T .

A direct consequence of property (ii) is that for almost every time $t \geq 0$ and all $f \in D$ the limit

$$\lim_{h \rightarrow 0} \frac{\langle f, \nu(t+h) \rangle - \langle f, \nu(t) \rangle}{h}$$

exists and is equal to $\langle f, \dot{\nu}(t) \rangle$. This justifies the notation $u(s) = \dot{\nu}(s)$.

Using these definitions, we are able to sharpen Theorem 2.1. In Section 4, we study the semigroup $V(t)$ and its generator H . Also, we give a number of properties of the level sets of the Lagrangian \mathcal{L} , defined in the theorem below. The proof of the theorem is given in Section 5.

Theorem 2.8. *Let (E, d) be locally compact. Let $(A, \mathcal{D}(A))$ have a core D equipped with a topology τ_D such that (D, τ_D) satisfies Condition 2.3. Then, the rate function in Theorem 2.1 can be rewritten as*

$$I(\nu) = \begin{cases} H(\nu(0) | \mathbb{P}_0) + \int_0^\infty \mathcal{L}(\nu(s), \dot{\nu}(s)) ds & \text{if } \nu \in \mathcal{AC}_{\nu(0)} \\ \infty & \text{otherwise,} \end{cases}$$

where $\mathcal{L} : \mathcal{P}(E) \times D' \rightarrow \mathbb{R}^+$ is given by

$$\mathcal{L}(\mu, u) := \sup_{f \in D} \{ \langle f, u \rangle - \langle Hf, \mu \rangle \}.$$

Remark 2.9. If we restrict ourselves to $[0, T]$ instead of \mathbb{R}^+ , then we obtain

$$\begin{aligned} & I^T(\{\nu(s)\}_{0 \leq s \leq T}) \\ &= \begin{cases} H(\nu(0) | \mathbb{P}_0) + \int_0^T \mathcal{L}(\nu(s), \dot{\nu}(s)) ds & \text{if } \nu \in \mathcal{AC}_{\nu(0)}^T \\ \infty & \text{otherwise,} \end{cases} \end{aligned}$$

by applying the contraction principle.

3 The large deviation principle via Sanov's theorem and optimal trajectories

Let (E, d) is a complete separable metric space. We start by proving the large deviation principle for a general class of processes via Sanov's theorem and the contraction principle. This will lead to the proof of Theorem 2.1.

Define for every t the map $\pi_t : D_E(\mathbb{R}^+) \rightarrow E$ by $\pi_t(x) := x(t)$. By Proposition III.7.1 in Ethier and Kurtz, π_t is a measurable map. Complementary to π_t , we introduce the map π_{t-} . For every path $x \in D_E(\mathbb{R}^+)$, the value $x(t-) := \lim_{s \uparrow t} x(s)$ is well defined, which makes it possible to define $\pi_{t-} : D_E(\mathbb{R}^+) \rightarrow E$ by $\pi_{t-}(x) := x(t-)$. As π_{t-} is the point-wise limit of the measurable maps π_{t_n} , for $t_n < t$, $t_n \uparrow t$, also π_{t-} is measurable.

Let \mathbb{P} be a probability measure on $D_E(\mathbb{R}^+)$, and let $X = (X(t))_{t \geq 0}$ be the process with law \mathbb{P} . Define $\mu(t) = \mathbb{P} \circ \pi_t^{-1}$ and $\mu(t-) = \mathbb{P} \circ \pi_{t-}^{-1}$ the laws of $X(t)$ and $X(t-)$. Also define the map $\phi : \mathcal{P}(D_E(\mathbb{R}^+)) \rightarrow \mathcal{P}(E)^{\mathbb{R}^+}$ by setting $\phi(\mathbb{P}) = (\mu(t))_{t \geq 0}$ and finally define the maps $\phi_t : \mathcal{P}(D_E(\mathbb{R}^+)) \rightarrow \mathcal{P}(E)$ by setting $\phi_t(\mathbb{P}) = \mu(t)$.

Lemma 3.1. *ϕ is a map from $\mathcal{P}(D_E(\mathbb{R}^+))$ to $D_{\mathcal{P}(E)}(\mathbb{R}^+)$.*

Proof. First, we prove that if $s \downarrow t$ then $\mu(s) \rightarrow \mu(t)$ weakly. Because the paths of X are right-continuous, we have $X(s) \rightarrow X(t)$. Hence, we have a.s. convergence, which in turn implies that $\mu(s) \rightarrow \mu(t)$ weakly.

If $s \uparrow t$, then we need to show that $\lim_{s \uparrow t} \mu(t)$ exists, but as above $X(s) \rightarrow X(t-)$, hence, the weak limit $\lim_{s \uparrow t} \mu(s)$ is equal to $\mu(t-)$. \square

We would like to prove that ϕ and $\{\phi_t\}_{t \geq 0}$ are continuous maps, but this is not always true as can be seen from the following example.

Example 3.2. Pick two distinct points e_1, e_2 in E . Define

$$x_n^+(t) = \begin{cases} e_1 & \text{for } t < 1 + 1/n \\ e_2 & \text{for } t \geq 1 + 1/n \end{cases}$$

$$x_n^-(t) = \begin{cases} e_1 & \text{for } t < 1 - 1/n \\ e_2 & \text{for } t \geq 1 - 1/n \end{cases}$$

and let $\mathbb{P}^n \in \mathcal{P}(D_E(\mathbb{R}^+))$ be defined by $\mathbb{P}^n = \frac{1}{2}\delta_{x_n^+} + \frac{1}{2}\delta_{x_n^-}$. Clearly, the sequence \mathbb{P}^n converges weakly to $\tilde{\mathbb{P}} = \delta_{\tilde{x}}$ where $\tilde{x}(t)$ is equal to e_1 for $t < 1$ and e_2 for $t \geq 1$.

If we look at the images $\phi(\mathbb{P}^n) = (\mu^n(t))_{t \geq 0}$ and $\phi(\tilde{\mathbb{P}}) = (\tilde{\mu}(t))_{t \geq 0}$, then we obtain

$$\mu^n(t) = \begin{cases} \delta_{e_1} & \text{for } t < 1 - 1/n \\ \frac{1}{2}\delta_{e_1} + \frac{1}{2}\delta_{e_2} & \text{for } 1 - 1/n \leq t < 1 + 1/n \\ \delta_{e_2} & \text{for } 1 + 1/n \leq t, \end{cases}$$

$$\tilde{\mu}(t) = \begin{cases} \delta_{e_1} & \text{for } t < 1 \\ \delta_{e_2} & \text{for } t \geq 1. \end{cases}$$

Clearly, $\mu^n(1) \rightarrow \frac{1}{2}\delta_{e_1} + \frac{1}{2}\delta_{e_2}$, which is not equal to $\tilde{\mu}(1)$ or $\tilde{\mu}(1-)$. We obtain that both ϕ and ϕ_1 are not continuous. Obviously, it follows that ϕ_t for $t \geq 0$ are not continuous either.

So problems arise when the time marginals of the limiting measure \mathbb{P} are discontinuous in time. However, this is the only thing that can happen.

Proposition 3.3. $\phi : \mathcal{P}(D_E(\mathbb{R}^+)) \rightarrow D_{\mathcal{P}(E)}(\mathbb{R}^+)$ is continuous at measures \mathbb{P} for which it holds that for every $t > 0$: $\mathbb{P}[X(t) = X(t-)] = 1$.

A similar statement for the finite dimensional projections ϕ_t , can be found in Ethier and Kurtz [12, Theorem 3.7.8].

Proof. Let $\mathbb{P}^n, \mathbb{P} \in \mathcal{P}(D_E(\mathbb{R}^+))$ such that $\mathbb{P}^n \rightarrow \mathbb{P}$ weakly and \mathbb{P} such that for every t $\mathbb{P}[X(t) = X(t-)] = 1$. By the Skorokhod representation Theorem [12, Theorem 3.1.9], we can find a probability space (Ω, \mathcal{F}, P) and $D_E(\mathbb{R}^+)$ valued random variables Y^n, Y distributed as X^n and X under \mathbb{P}^n, \mathbb{P} such that $Y^n \rightarrow Y$ P a.s.

Let $\{t_n\}_{n \geq 0}$ be a sequence converging to $t > 0$. Define the sets

$$\begin{aligned} A &:= \{Y(t) = Y(t-)\}, \\ B &:= \{d(Y^n(t_n), Y(t)) \wedge d(Y^n(t_n), Y(t-)) \rightarrow 0\}. \end{aligned}$$

By the assumption that $\mathbb{P}[X(t) = X(t-)] = 1$, it follows that $P[A] = 1$. By Proposition 3.6.5 in [12], and the fact that $Y^n \rightarrow Y$ P a.s. it follows that $P[B] = 1$. Combining these statements yields

$$P[Y^n(t_n) \rightarrow Y(t)] \geq P[A \cap B] = 1,$$

which implies that $\mu^n(t_n) \rightarrow \mu(t)$. As $\mu(t) = \mu(t-)$ by assumption, Proposition 3.6.5 in Ethier and Kurtz yields the final result. \square

3.1 Large deviations for measures on the Skorokhod space

Suppose that we have a process X on $D_E(\mathbb{R}^+)$ and a corresponding measure $\mathbb{P} \in \mathcal{P}(D_E(\mathbb{R}^+))$. Then Sanov's theorem, Theorem 6.2.10 in [8], gives us the large deviation behaviour of the empirical distribution L_n^X of independent copies of the process X : X^1, X^2, \dots :

$$L_n^X := \frac{1}{n} \sum_{i=1}^n \delta_{\{X^i\}} \in \mathcal{P}(D_E(\mathbb{R}^+)).$$

Theorem 3.4 (Sanov). *The empirical measures L_n^X satisfy the large deviation principle on $\mathcal{P}(D_E(\mathbb{R}^+))$ with respect to the weak topology with the good and convex rate function*

$$I^*(\mathbb{Q}) = H(\mathbb{Q} \mid \mathbb{P}) := \int \frac{d\mathbb{Q}}{d\mathbb{P}} \log \frac{d\mathbb{Q}}{d\mathbb{P}} d\mathbb{P}.$$

We are interested in obtaining a large deviation principle on $D_{\mathcal{P}(E)}(\mathbb{R}^+)$. In Proposition 3.3, we saw that we have a map ϕ that is continuous on a part of its domain. Hence, we are in the position to use the contraction principle.

Theorem 3.5. *Suppose that \mathbb{P} satisfies $\mathbb{P}[X(t) = X(t-)] = 1$ for every $t \geq 0$, then the large deviation principle holds for*

$$\left(L_n^{X(t)}\right)_{t \geq 0} = \left(\frac{1}{n} \sum_{i=1}^n \delta_{X^i(t)}\right)_{t \geq 0}$$

on $D_{\mathcal{P}(E)}(\mathbb{R}^+)$ with rate function

$$I((\nu_t)_{t \geq 0}) = \inf\{H(\mathbb{Q} | \mathbb{P}) \mid \mathbb{Q} \in \mathcal{P}(D_E(\mathbb{R}^+)), \phi(\mathbb{Q}) = (\nu(t))_{t \geq 0}\}$$

and I is finite only on $C_{\mathcal{P}(E)}(\mathbb{R}^+)$.

Proof. The measures \mathbb{Q} for which it holds that $I(\mathbb{Q}) < \infty$ satisfy $\mathbb{Q} \ll \mathbb{P}$ hence it follows that for every t : $\mathbb{Q}[X(t) = X(t-)] = 1$. This yields that ϕ is continuous at \mathbb{Q} by Proposition 3.3.

By the contraction principle, Theorem 4.2.1 and remark (c) after Theorem 4.2.1 in Dembo and Zeitouni [8], we obtain the large deviation principle on $D_{\mathcal{P}(E)}(\mathbb{R}^+)$ with I as given in the theorem. \square

3.2 The large deviation principle for Markov processes

Although Theorem 3.5 can be applied to a wide range of (time-inhomogeneous) processes, we explore its consequences for time-homogeneous Markov processes. Recall the definition of a solution to the martingale problem preceding Theorem 2.1.

Lemma 3.6. *Suppose that the process X with corresponding measure \mathbb{P} on $D_E(\mathbb{R}^+)$ solves the martingale problem for $(A, \mathcal{D}(A))$ with starting measure \mathbb{P}_0 . Then, it holds that for every $t \geq 0$ $\mathbb{P}[X(t) = X(t-)] = 1$. Hence, the large deviation principle holds for $\{L_n^{X(t)}\}_{t \geq 0}$ on $D_{\mathcal{P}(E)}(\mathbb{R}^+)$ with rate function*

$$I((\nu_t)_{t \geq 0}) = \inf\{H(\mathbb{Q} | \mathbb{P}) \mid \mathbb{Q} \in \mathcal{P}(D_E(\mathbb{R}^+)), \phi(\mathbb{Q}) = (\nu(t))_{t \geq 0}\}$$

and I is finite only on $C_{\mathcal{P}(E)}(\mathbb{R}^+)$.

Proof. To apply Theorem 3.5, we need to check that $\mathbb{P}[X(t) = X(t-)] = 1$ for every $t \geq 0$, but this follows by Theorem 4.3.12 in [12]. \square

Using this result, Theorem 2.1 follows without much effort.

Proof of Theorem 2.1. The large deviation principle follows by Lemma 3.6. This lemma also gives that the rate function is ∞ on the complement of $C_{\mathcal{P}(E)}(\mathbb{R}^+)$. To obtain the rate function as a supremum over rate functions for finite dimensional problems

$$I(\nu) = \begin{cases} \sup_{0, t_1, \dots, t_k} I[0, t_1, \dots, t_k](\nu(0), \nu(t_1), \dots, \nu(t_k)) & \text{if } \nu \in C_{\mathcal{P}(E)}(\mathbb{R}^+) \\ \infty & \text{otherwise,} \end{cases}$$

we use Theorem 4.13 and Theorem 4.30 in Feng and Kurtz [13]. Proposition 7.3 gives us the final decomposition of the rate function. \square

3.3 Approximating $V(t)$

Before we turn to the proof of Theorem 2.8, we start with some general results on approximating $\langle V(t)f, \mu \rangle = \sup_{\nu \in \mathcal{P}(E)} \langle f, \nu \rangle - I_t(\nu | \mu)$ by using the Doob transform. For related results on Schrödinger bridges, see Föllmer and Gantert [17] or the survey paper by Léonard [22].

First, we introduce the Doob transform, see Doob [11, page 566] or [17, 18]. Fix a Markov measure $\mathbb{P} \in \mathcal{P}(D_E(\mathbb{R}^+))$ corresponding to a semigroup of transition operators $\{S(t)\}_{t \geq 0}$, with transition probabilities $P_{s,t}(x, dy)$ and define the measure $\mathbb{Q} \in \mathcal{P}(D_E(\mathbb{R}^+))$ by

$$\frac{d\mathbb{Q}}{d\mathbb{P}}(X) = \rho_0(X(0))\rho_1(X(t)),$$

where ρ_0, ρ_1 are measurable and non-negative functions on E such that

$$\int \rho_0(X(0))\rho_1(X(t))\mathbb{P}(dX) = 1.$$

Define the non-negative and measurable function $k : E \times [0, t] \rightarrow \mathbb{R}$ by

$$k(x, s) = \int \rho_1(y)P_{s,t}(x, dy). \quad (3.1)$$

k is space-time harmonic: for $0 \leq r \leq s \leq t$, we have

$$k(x, r) = \int P_{r,s}(x, dy)k(y, s). \quad (3.2)$$

The measure \mathbb{Q} is Markov, the law \mathbb{Q}_0 at time 0 is given by $\frac{d\mathbb{Q}_0}{d\mathbb{P}_0}(X(0)) = \rho_0(X(0))k(X(0), 0)$ and the transition probabilities $Q_{r,s}$ are given by

$$Q_{s,r}(x, dy) = k(x, r)^{-1}P_{r,s}(x, dy)k(y, s)$$

for $s, r \leq t$ and $Q_{s,r}(x, dy) = P_{s,r}(x, dy)$ for $s, r \geq t$.

Proposition 3.7. *Let $\mu \in \mathcal{P}(E)$ and let $\mathbb{P} \in \mathcal{P}(D_E(\mathbb{R}^+))$ be the Markov measure with initial law $\mathbb{P}_0 = \mu$ and transition semigroup $\{S(t)\}_{t \geq 0}$. Let $f \in C_b(E)$ and $t \geq 0$. Consider the optimisation problem*

$$\langle V(t)f, \mu \rangle = \sup_{\nu \in \mathcal{P}(E)} \{ \langle f, \nu \rangle - I_t(\nu | \mu) \}, \quad (3.3)$$

where I_t is defined as in Equation (2.1).

Then, there exists a Markov measure $\mathbb{Q}^* \ll \mathbb{P}$ such that the law $\phi_t(\mathbb{Q})$ of $X(t)$ under \mathbb{Q} optimises the supremum in (3.3), $I_t(\phi_t(\mathbb{Q}^*) | \mathbb{P}_0) = H(\mathbb{Q}^* | \mathbb{P})$ and there are Markov measures \mathbb{Q}^n with the following properties:

- (a) $H(\mathbb{Q}^n | \mathbb{Q}^*) \rightarrow 0$, and as a consequence $H(\mathbb{Q}_0^n | \mathbb{P}_0) \rightarrow 0$ and $\|\mathbb{Q}^n - \mathbb{Q}^*\|_{TV} \rightarrow 0$.
- (b) $H(\mathbb{Q}^n | \mathbb{P}) \uparrow H(\mathbb{Q}^* | \mathbb{P})$.
- (c) \mathbb{Q}_n is a Doob transform of \mathbb{P} , i.e. there are non-negative measurable functions ρ_n and constants λ_n such that

$$\frac{d\mathbb{Q}^n}{d\mathbb{P}}(X) = \rho_n(X(0)) \exp \{ \lambda_n f(X(t)) \}.$$

The proof is inspired by the proof of Theorem 3.43 in [17].

Proof. We denote with $\mathbb{P}_0 \in \mathcal{P}(E)$ and $\mathbb{P}_{0,t}$ the restriction of \mathbb{P} to the time 0, respectively, time 0 and time t coordinates. First, note that as $\mu = \mathbb{P}_0$, we have

$$\begin{aligned} I_t(\nu | \mu) &= H(\mu | \mu) + I_t(\nu | \mu) \\ &= \inf_{\substack{\mathbb{Q} \in \mathcal{P}(D_E(\mathbb{R}^+)) \\ \phi_0(\mathbb{Q}) = \mu, \phi_t(\mathbb{Q}) = \nu}} H(\mathbb{Q} | \mathbb{P}) \end{aligned}$$

by the contraction principle. We obtain

$$\begin{aligned} \langle V(t)f, \mu \rangle &= \sup_{\nu \in \mathcal{P}(E)} \{ \langle f, \nu \rangle - I_t(\nu | \mu) \} \\ &= \sup_{\nu \in \mathcal{P}(E)} \sup_{\substack{\mathbb{Q} \in \mathcal{P}(D_E(\mathbb{R}^+)) \\ \phi_0(\mathbb{Q}) = \mu, \phi_t(\mathbb{Q}) = \nu}} \{ \langle f, \phi_t(\mathbb{Q}) \rangle - H(\mathbb{Q} | \mathbb{P}) \} \\ &= \sup_{\substack{\mathbb{Q} \in \mathcal{P}(D_E(\mathbb{R}^+)) \\ \phi_0(\mathbb{Q}) = \mu}} \{ \langle f, \phi_t(\mathbb{Q}) \rangle - H(\mathbb{Q} | \mathbb{P}_0) \}. \end{aligned}$$

As the relative entropy has compact level sets, there is an optimising measure $\mathbb{Q}^* \in \mathcal{P}(D_E(\mathbb{R}^+))$. Set $\alpha_0 = \langle f, \phi_t(\mathbb{Q}^*) \rangle$. By choice of α_0 , we obtain

$$\begin{aligned} \langle V(t)f, \mu \rangle &= \sup_{\substack{\mathbb{Q} \in \mathcal{P}(D_E(\mathbb{R}^+)) \\ \phi_0(\mathbb{Q}) = \mu}} \{ \langle f, \phi_t(\mathbb{Q}) \rangle - H(\mathbb{Q} | \mathbb{P}) \} \\ &= \sup_{\substack{\mathbb{Q} \in \mathcal{P}(D_E(\mathbb{R}^+)) \\ \phi_0(\mathbb{Q}) = \mu, \langle f, \phi_t(\mathbb{Q}) \rangle = \alpha_0}} \{ \langle f, \phi_t(\mathbb{Q}) \rangle - H(\mathbb{Q} | \mathbb{P}) \}. \end{aligned}$$

E is Polish which implies that the closed set $\{\mathbb{Q} \in \mathcal{P}(D_E(\mathbb{R}^+)) \mid \phi_0(\mathbb{Q}) = \mu\}$ can be written as $\{\mathbb{Q} \in \mathcal{P}(D_E(\mathbb{R}^+)) \mid \langle g_i, \phi_0(\mathbb{Q}) \rangle = \alpha_i, i \geq 1\}$ for some countable set of functions $\{g_i\}_{i \geq 1}$ in $C_b(E)$ and real numbers $\alpha_i = \int g_i d\mu$, $i \geq 1$. For $n \in \mathbb{N}$, define the approximating convex and weakly closed sets

$$S_n := \{\mathbb{Q} \in \mathcal{P}(D_E(\mathbb{R}^+)) \mid \langle f, \phi_t(\mathbb{Q}) \rangle = \alpha_0, \langle g_i, \phi_0(\mathbb{Q}) \rangle = \alpha_i, 1 \leq i \leq n\}.$$

For each $n \in \mathbb{N}$, there exists a unique measure $\mathbb{Q}^n \in S_n$, see Csiszár [6], such that

$$H(\mathbb{Q}^n | \mathbb{P}) = \inf_{\mathbb{Q} \in S_n} H(\mathbb{Q} | \mathbb{P}),$$

which has a density ψ^n with respect to \mathbb{P} of the form

$$\psi^n(x, y) = Z_n^{-1} \exp \left\{ \lambda_{n,0} f(y) + \sum_{k=1}^n \lambda_{n,k} g_k(x) \right\}$$

with constants Z_n and $\lambda_{n,k}$, $n \in \mathbb{N}$, $0 \leq k \leq n$.

Lemma 1.23 in Föllmer [16, Section II.1.3, Page 163] gives us $H(\mathbb{Q}^n | \mathbb{Q}^*) \rightarrow 0$ and by contracting to the first coordinate $H(\phi_0(\mathbb{Q}^n) | \mu) \rightarrow 0$. The inequality $\|\alpha - \beta\|^2 \leq 2H(\alpha | \beta)$ for arbitrary probability measures α, β yields

$$\lim_{n \rightarrow \infty} \|\mathbb{Q}^n - \mathbb{Q}^*\|_{TV} = 0. \quad (3.4)$$

Additionally, Lemma 1.23 [16] yields

$$\lim_{n \rightarrow \infty} H(\mathbb{Q}^n | \mathbb{P}) = H(\mathbb{Q}^* | \mathbb{P}). \quad (3.5)$$

Together with the fact that the sets S_n are decreasing, (b) follows. By construction the measures \mathbb{Q}^n are Markov. As convergence in total variation preserves the Markov property, also \mathbb{Q} is Markov. \square

4 A study of the operators $V(t)$, H , L and A^g .

4.1 The semigroup $V(t)$ and the generator H

We return to the situation that (E, d) is a locally compact metric space, so that we can use semigroup theory to rewrite the rate function.

First suppose that E is non-compact. Let $E^\Delta = E \cup \{\Delta\}$ be the one-point compactification. By Lemma 4.3.2 in [12], $S(t)$ extends to a strongly continuous contraction semigroup on $(C(E^\Delta), \|\cdot\|)$ by setting $S^\Delta(t)f = f(\Delta) + S(t)(f - f(\Delta))$. Therefore, we can argue using the semigroup on the compact space E^Δ , and then obtain the result in Theorem 2.8 on E by Theorem 4.11 in Feng and Kurtz [13].

From this point onward, we assume that (E, d) is compact and that the transition semigroup $\{S(t)\}_{t \geq 0}$ is strongly continuous on $C(E)$. Let $A : \mathcal{D}(A) \subseteq C(E) \rightarrow C(E)$ be the associated infinitesimal generator.

We examine $V(t)f(x) = \log S(t)e^f(x) = \log \mathbb{E}[e^{f(X(t))} \mid X(0) = x]$, $f \in C(E)$, which was defined in Theorem 2.1. It is an elementary calculation to check that $\{V(t)\}_{t \geq 0}$ is a strongly continuous contraction semigroup on $C(E)$.

As in the linear case, we try to calculate the generator of V :

$$Hf = \lim_{t \downarrow 0} \frac{V(t)f - f}{t}$$

defined for $f \in \mathcal{D}(H)$, where

$$\mathcal{D}(H) := \left\{ f \in C(E) \mid \exists g \in C(E) : \lim_{t \downarrow 0} \left\| \frac{V(t)f - f}{t} - g \right\| = 0 \right\}.$$

We start with an extension of the chain rule to Banach spaces. The proof is rather standard and is left to the reader.

Lemma 4.1. *Let $f \in \mathcal{D}(A)$ and let $\phi : f(E) \rightarrow \mathbb{R}$ be differentiable on $f(E)$ and let ϕ' be Lipschitz continuous. Then it holds that*

$$\frac{d}{dt} \phi(S(t)f)|_{t=0} = \phi'(f)Af,$$

which should be interpreted as

$$\lim_{t \rightarrow 0} \frac{\phi(S(t)f) - \phi(f)}{t} = \phi'(f)Af$$

with respect to the sup norm.

A direct consequence is that we can calculate the generator H of $V(t)$ on a subset of its domain.

Corollary 4.2. *For $f \in C(E)$ such that $e^f \in \mathcal{D}(A)$, we have $f \in \mathcal{D}(H)$ and*

$$Hf = e^{-f} A(e^f).$$

In order to proceed, we need Condition 2.2. We see that Corollary 4.2 gives us that if $f \in \mathcal{D}$, then $f \in \mathcal{D}(H)$ and $Hf = e^{-f} A(e^f)$.

We will use this operator (H, D) , under Condition 2.3, to construct a new *Nisio* semigroup $\{\mathbf{V}(t)\}_{t \geq 0}$ on $C(\mathcal{P}(E))$. This semigroup will be introduced in Section 5.2, and there we will show that for $\mu \in \mathcal{P}(E)$ and $f \in C(E)$, we have $\mathbf{V}(t)[f](\mu) = \langle V(t)f, \mu \rangle$, where $[f] \in C(\mathcal{P}(E))$ is the function defined by $[f](\mu) = \langle f, \mu \rangle$.

We start with some results on $V(t)f$ and H that will be useful for proving the equality $\mathbf{V}(t)[f](\mu) = \langle V(t)f, \mu \rangle$. For $f \in C(E)$, let $J(\lambda)f := (\mathbb{1} - \lambda A)^{-1}f = \int_0^\infty \lambda^{-1} e^{-\lambda^{-1}t} S(t)f dt$. Using $J(\lambda)$, we set $R(\lambda)f := \log J(\lambda)e^f$.

We constructed the semigroup $V(t)$ from the linear semigroup $S(t)$, and the operator $R(\lambda)$ from the linear resolvent $J(\lambda)$. One would therefore hope that $R(\lambda)$ equals $(\mathbb{1} - \lambda H)^{-1}$. This is not the case, but we do have the following two results, which we will need for the proof of Lemma 5.8 and Proposition 5.10.

Lemma 4.3. *For $f \in C(E)$, we have $R(\lambda)f \in \mathcal{D}(H)$ and $(\mathbb{1} - \lambda H)R(\lambda)f \geq f$.*

Proof. $J(\lambda)$ maps $C(E)$ bijectively on $\mathcal{D}(A)$, therefore, $e^{R(\lambda)f} = J(\lambda)e^f \in \mathcal{D}(A)$. Thus by Corollary 4.2, we have that $R(\lambda)f \in \mathcal{D}(H)$.

Let $x \in E$, we prove $(\mathbb{1} - \lambda H)R(\lambda)f(x) \geq f(x)$. First, we show that We prove that the following quantity is larger than 0:

$$\begin{aligned} (\mathbb{1} - \lambda H)R(\lambda)f(x) - f(x) &= R(\lambda)f(x) - f(x) - \lambda \frac{AJ(\lambda)e^f(x)}{J(\lambda)e^f(x)} \\ &= R(\lambda)f(x) - f(x) - \frac{J(\lambda)e^f(x) - e^{f(x)}}{J(\lambda)e^f(x)}. \end{aligned}$$

This is equivalent to showing that

$$J(\lambda)e^f(x) \log(J(\lambda)e^f(x)) - f(x)J(\lambda)e^f(x) - J(\lambda)e^f(x) + e^{f(x)}$$

is positive, which follows from the fact that for every $c \in \mathbb{R}$, the function defined for non-negative y , given by $y \mapsto y \log y - (c+1)y + e^c$ is non-negative. \square

Note that the fact that the function $y \mapsto y \log y - (c+1)y + e^c$ has a unique point where it hits 0. This means that $(\mathbb{1} - \lambda H)R(\lambda)f(x) = f(x)$ only if $\mathbb{E}[e^{f(X_\tau)} | X_0 = x] = e^{f(x)}$, where τ is an exponential random variable with mean λ independent of the process X . This can not be true in general.

Even though $R(\lambda)$ does not invert $(\mathbb{1} - \lambda H)$, it does approximate the semigroup in a way that the resolvents of H would as well.

Lemma 4.4. *For every $f \in C(E)$, we have that $\lim_{n \rightarrow \infty} R(n^{-1})^{\lfloor nt \rfloor} f = V(t)f$.*

Proof. By definition, we have $R(n^{-1})^{\lfloor nt \rfloor} f = \log J(n^{-1})^{\lfloor nt \rfloor} e^f$. For linear semigroups, we know that the resolvents approximate the semigroup: $J(\frac{1}{n})^{\lfloor nt \rfloor} e^f \rightarrow S(t)e^f$, see for example Corollary 1.6.8 in [12]. Therefore, by uniform continuity of the logarithm on $[e^{-\|f\|}, e^{\|f\|}]$, we obtain the final result by applying the logarithm. \square

4.2 Operator duality for H

Additionally to the operator H , we introduce operators A^g that serve as generators of tilted Markov processes obtained from $X(t)$ by the change of measure given in Equation (2.4). We also introduce an operator L , that will serve as a precursor to our final Lagrangian \mathcal{L} .

Definition 4.5. Under Condition 2.2, define the following operators for $f, g \in D$:

$$\begin{aligned} Hf &= e^{-f} A e^f, \\ A^g f &= e^{-g} A (f e^g) - (e^{-g} A e^g) f, \\ Lg &= A^g g - Hg. \end{aligned}$$

H will be called the Hamiltonian and L the (pre-)Lagrangian in analogy to the Lagrangian and Hamiltonian of classical mechanics. A^g is a generator itself, see for example Palmowski and Rolski [25]. This is also illustrated by the next two examples.

We calculate H and A^g in the case of a Markov jump process and a standard Brownian motion.

Example 4.6. Let E be a finite set and let $\{X(t)\}_{t \geq 0}$ be generated by

$$Af(x) = \sum_y r(x, y) [f(y) - f(x)],$$

where r is some transition kernel. A calculations shows that

$$\begin{aligned} Hf(x) &= \sum_y r(x, y) [e^{f(y)-f(x)} - 1], \\ A^g f(x) &= \sum_y r(x, y) e^{g(y)-g(x)} [f(y) - f(x)]. \end{aligned}$$

Example 4.7. Let $E = \mathbb{R}$, and let $\{X(t)\}_{t \geq 0}$ be a standard Brownian motion, for which the generator A is given for $f \in C_c^\infty(\mathbb{R})$, i.e. smooth and compactly supported functions, by $Af(x) = \frac{1}{2}f''(x)$. H and A^g are given by

$$\begin{aligned} Hf(x) &= \frac{1}{2}f''(x) + \frac{1}{2}(f'(x))^2, \\ A^g f(x) &= \frac{1}{2}f''(x) + f'(x)g'(x). \end{aligned}$$

In both examples, it is seen that A^g is also a generator of a Markov process. More importantly, however, L and H are operator duals.

Lemma 4.8. Under Condition 2.2, we have for $f \in D$ that

$$\langle Hf, \mu \rangle = \sup_{g \in D} \{ \langle A^g f, \mu \rangle - \langle Lg, \mu \rangle \}, \quad (4.1)$$

and equality holds for $g = f$. Furthermore, for $g \in D$ and $\mu \in \mathcal{P}(E)$ it holds that

$$\langle Lg, \mu \rangle = \sup_{f \in D} \{ \langle A^g f, \mu \rangle - \langle Hf, \mu \rangle \}, \quad (4.2)$$

with equality for $f = g$.

Proof. For $\lambda > 0$, let $A_\lambda := \lambda^{-1}(J(\lambda) - \mathbb{1})$ be the Yosida approximant of A . It is well known that A_λ is bounded and is given by

$$A_\lambda f(x) = \lambda^{-1} \int q_\lambda(x, dy) [f(y) - f(x)],$$

where $q_\lambda(x, \cdot)$ is the law of the process generated by A after an exponential random time with mean λ .

Now define H_λ, A_λ^g and L_λ in terms of A_λ . As A_λ is bounded, it follows by Lemma 5.7 in [13] that

$$\begin{aligned} H_\lambda f(x) &\geq A_\lambda^g f(x) - L_\lambda g(x), \\ H_\lambda f(x) &= A_\lambda^f f(x) - L_\lambda f(x). \end{aligned}$$

Therefore, it follows by Yosida approximation [12, Lemma 1.2.4] that

$$Hf(x) = \sup_{g \in D} \{A^g f(x) - Lg(x)\}.$$

The first statement now follows by integration. The variational statement for L is obtained similarly. \square

4.3 The Lagrangian and a variational expression for the Hamiltonian

The Lagrangian in the previous section is still an operator acting on functions. Here we embed this object in a new Lagrangian \mathcal{L} that is a function of place and speed. Also, we introduce a map ρ that transforms ‘momentum’ into speed.

Definition 4.9. Let (D, τ_D) satisfy Condition 2.3. Define the Lagrangian $\mathcal{L} : \mathcal{P}(E) \times D' \rightarrow \mathbb{R}^+$ by

$$\mathcal{L}(\mu, u) = \sup_{f \in D} \{\langle f, u \rangle - \langle Hf, \mu \rangle\}.$$

Also, define the map $\rho : \mathcal{P}(E) \times D \rightarrow D'$ by $\rho(\mu, g) = (A^g)'(\mu)$.

\mathcal{L} can be considered as an extension of L . Pick $\mu \in \mathcal{P}(E)$ and $g \in D$, then

$$\begin{aligned} \mathcal{L}(\mu, \rho(\mu, g)) &= \sup_{f \in D} \{\langle f, \rho(\mu, g) \rangle - \langle Hf, \mu \rangle\} \\ &= \sup_{f \in D} \{\langle A^g f, \mu \rangle - \langle Hf, \mu \rangle\} \\ &= \langle Lg, \mu \rangle, \end{aligned} \tag{4.3}$$

where the last equality follows by Equation (4.2).

Lemma 4.10. $(\mu, u) \mapsto \mathcal{L}(\mu, u)$ is convex and lower semi-continuous with respect to the weak and weak* topologies.

Proof. \mathcal{L} is lower semi-continuous, because it is the supremum over continuous functions. Convexity of \mathcal{L} follows by the linearity of $u \mapsto \langle f, u \rangle$ and $\mu \mapsto \langle Hf, \mu \rangle$. \square

It turns out that the space D' is too large for practical purposes. Recall the set \mathcal{N} introduced in Condition 2.3 (f) and the definition of a Polar in (2.2). Define $U \subseteq D'$ by

$$U := \bigcup_{n \in \mathbb{N}} n\mathcal{N}^\circ. \quad (4.4)$$

We equip U with the weak* topology inherited from D' . The importance of U follows from the following lemma, which shows that we can restrict the set of allowed ‘speeds’ to U .

Lemma 4.11. *Let $\mu \in \mathcal{P}(E)$. If $u \notin U$, then $\mathcal{L}(\mu, u) = \infty$. Furthermore, for $\mu \in \mathcal{P}(E)$, we have $\rho(\mu, g) \in U$.*

Proof. For $u \notin U = \bigcup_n n\mathcal{N}^\circ$, we can find functions $f_n \in \mathcal{N}$, such that $|\langle f_n, u \rangle| \geq n$. The inequality $|\langle f_n, u \rangle| \leq \mathcal{L}(\mu, u) + \langle Hf_n, \mu \rangle \vee \langle H(-f_n), \mu \rangle$, yields that $\mathcal{L}(\mu, u) \geq n - 1$ for every n , which implies that $\mathcal{L}(\mu, u) = \infty$.

The second statement follows from the first, Equation (4.3), and the fact that Lg is bounded. \square

As can be seen from Equation (4.3), \mathcal{L} is an extension of L . As expected, H can also be obtained by a Fenchel-Legendre transform of \mathcal{L} .

Lemma 4.12. *The variational expression for H in Equation (4.1) extends to*

$$\begin{aligned} \langle Hf, \mu \rangle &= \sup_{u \in D'} \{ \langle f, u \rangle - \mathcal{L}(\mu, u) \} \\ &= \sup_{u \in U} \{ \langle f, u \rangle - \mathcal{L}(\mu, u) \} \end{aligned}$$

Proof. First of all, note the equality of the two variational expressions on the right hand side, as $\mathcal{L}(\mu, u) = \infty$ if $u \notin U$. We give two proofs of the first equality. First of all, Hölder's inequality tells us that $f \mapsto \langle V(t)f, \mu \rangle$ is convex. Therefore, Hf is the norm, and thus, point-wise limit of convex functions which implies that $f \mapsto \langle Hf, \mu \rangle$ is convex. The result follows directly from the fact that the double Fenchel-Legendre transform of the convex lower semi-continuous function $f \mapsto \langle Hf, \mu \rangle$ is $f \mapsto \langle Hf, \mu \rangle$ by the Fenchel-Moreau theorem [8, Lemma 4.5.8]. The second approach is more direct. By Definition 4.9 of \mathcal{L} , we obtain that for every $f \in D$, $\mu \in \mathcal{P}(E)$, $u \in D'$: $\langle Hf, \mu \rangle \geq \langle f, u \rangle - \mathcal{L}(\mu, u)$.

We now show that we in fact have equality. By Equation (4.3), we know that $\mathcal{L}(\mu, \rho(\mu, g)) = \langle Lg, \mu \rangle$. Hence, by the second item in Lemma 4.8, we obtain

$$\begin{aligned} \langle Hf, \mu \rangle &= \langle A^f f, \mu \rangle - \langle Lf, \mu \rangle \\ &= \langle f, \rho(\mu, f) \rangle - \mathcal{L}(\mu, \rho(\mu, f)), \end{aligned} \quad (4.5)$$

which concludes the proof. \square

The latter approach in the proof of Lemma 4.12, gives us even more information.

Proposition 4.13. *Let $\mu \in \mathcal{P}(E)$ and define Γ_μ to be the weak* closed convex hull of $\{\rho(\mu, g) \in U \mid g \in D\}$. If $u \notin \Gamma_\mu$, then $\mathcal{L}(\mu, u) = \infty$.*

Proof. Fix $\mu \in \mathcal{P}(E)$. Define $\hat{\mathcal{L}} = \mathcal{L}$ if $u \in \Gamma_\mu$ and set $\hat{\mathcal{L}} = \infty$ for $u \notin \Gamma_\mu$. It is clear that $\hat{\mathcal{L}}$ is also convex and lower semi-continuous. As $\hat{\mathcal{L}} \geq \mathcal{L}$, it follows by Lemma 4.12 that $\langle Hf, \mu \rangle \geq \sup_u \{ \langle f, u \rangle - \hat{\mathcal{L}}(\mu, u) \}$.

As in Equation (4.5), we obtain

$$\begin{aligned}\langle Hf, \mu \rangle &= \langle A^f f, \mu \rangle - \langle Lf, \mu \rangle \\ &= \langle f, \rho(\mu, f) \rangle - \mathcal{L}(\mu, \rho(\mu, f)) \\ &= \langle f, \rho(\mu, f) \rangle - \hat{\mathcal{L}}(\mu, \rho(\mu, f)),\end{aligned}$$

which shows that $\langle Hf, \mu \rangle = \sup_u \{ \langle f, u \rangle - \hat{\mathcal{L}}(\mu, u) \}$. In other words, the double Fenchel-Legendre transform of the convex and lower semi-continuous function $\hat{\mathcal{L}}$ is \mathcal{L} . This implies that they are equal. \square

4.4 The Doob-h transform in terms of tilted generators

We connect the operators introduced in the last few sections to the discussion on the Doob-transform in Section 3.3. There, we considered a measure $\mathbb{Q} \in \mathcal{P}(D_E(\mathbb{R}^+))$, defined by

$$\frac{d\mathbb{Q}}{d\mathbb{P}}(X) = \rho_0(X(0))\rho_1(X(t)).$$

As a consequence of the result of Proposition 3.7, we will switch our focus to measures of the type

$$\frac{d\mathbb{Q}}{d\mathbb{P}}(X) = \rho_0(X(0)) \exp \{h(X(t))\} = \frac{d\mathbb{Q}_0}{d\mathbb{P}_0}(X(0)) \exp \{h(X(t)) - h(0)(X(0))\} \quad (4.6)$$

where now we define $h(s) = V(s, t)h$, or $e^{h(s)} = S(s, t)e^h$. This is just the definition of the Doob-transform as in (3.1), where $e^h = k$. The transition probabilities of the Markov process described by \mathbb{Q} up to time t can also be restated as a semigroup of transition operators $\{S^{h[0, t]}(r, s)\}_{0 \leq r \leq s \leq t}$, where $S^{h[0, t]}(r, s) : C(E) \rightarrow C(E)$ is defined by $S^{h[0, t]}(r, s)f(x) := \mathbb{Q}[f(X(s)) \mid X(r) = x]$, which is obtained from Equation (3.2) and given by

$$S^{h[0, t]}(r, s)f(x) = e^{-h(r)}(x)S(r, s) \left(f e^{h(s)} \right) (x).$$

It is straightforward to check that $(r, s) \mapsto S^{h[0, t]}(r, s)f$ is continuous for all $f \in C(E)$ and $(r, s) \in \{(r', s') \mid 0 \leq r' \leq s' \leq t\}$. We can say more even. The next lemma shows that the tilted generators of the previous section turn up in the study of this semigroup. After that, we will show that $H(\mathbb{Q} \mid \mathbb{P})$ can be given in terms of an integral over the Lagrangian \mathcal{L} . We start with two definitions.

Let $C([0, t], D)$ be the space of trajectories $\{g(s)\}_{s \in [0, t]}$, $g(s) \in D$ such that $s \mapsto g(s)$ is continuous with respect to τ_D . Furthermore, let $C^1([0, t], D) \subseteq C([0, t], D)$ be those trajectories for which there exists a trajectory $\{\partial g(s)\}_{s \in [0, t]}$ in $C([0, t], C(E))$ such that for all $s \in [0, t]$, we have

$$\lim_{r \rightarrow 0} \left\| \frac{g(s+r) - g(s)}{r} - \partial g(s) \right\| = 0.$$

Now suppose that $h \in D$, then Condition 2.2 (b) and (c) imply that $h(s) = V(t-s)h \in D$ for all $s \in [0, t]$. In this case, we can find the trajectory of generators of the semigroup $S^{h[0, t]}$.

Proposition 4.14. Suppose that $h \in D$. For every $s \in [0, t]$ and $f \in D$, we have

$$\lim_{r \rightarrow 0} \frac{S^{h[0,t]}(s, s+r)f - f}{r} = A^{h(s)}f.$$

If $f \in C^1([0, t], D)$, then we have for every $s \in [0, t]$ that

$$\lim_{r \rightarrow 0} \frac{S^{h[0,t]}(s, s+r)f(s+r) - f(s)}{r} = A^{h(s)}f(s) + \partial f(s).$$

Proof. We start with the proof of the first statement. Let $f \in D$ and $s \in [0, t]$, we prove the result for $r > 0$, the proof of the other side is similar. Clearly,

$$\lim_{r \rightarrow 0} \left\| \frac{S^{h[0,t]}(s, s+r)f - f}{r} - A^{h(s)}f \right\| = 0$$

if and only if

$$\lim_{r \rightarrow 0} \left\| e^{h(s)} \left[\frac{S^{h[0,t]}(s, s+r)f - f}{r} - A^{h(s)}f \right] \right\| = 0.$$

Therefore, we will prove the latter. We see

$$\begin{aligned} & e^{h(s)} \left[\frac{S^{h[0,t]}(s, s+r)f - f}{r} - A^{h(s)}f \right] \\ &= e^{h(s)} \left[\frac{e^{-h(s)}S(r)(fe^{h(s+r)}) - f}{r} - A^{h(s)}f \right] \\ &= \frac{S(r)(fe^{h(s+r)}) - e^{h(s)}f}{r} - A(fe^{h(s)}) + fAe^{h(s)} \\ &= \frac{S(r)(fe^{h(s+r)}) - S(r)(fe^{h(s)})}{r} + S(r)(fAe^{h(s)}) \\ &\quad + \frac{S(r)(fe^{h(s)}) - fe^{h(s)}}{r} - A(fe^{h(s)}) \\ &\quad + fAe^{h(s)} - S(r)(fAe^{h(s)}). \end{aligned}$$

The last two lines converge to 0 as $r \downarrow 0$. We consider the term in line four. First note that $S(r)$ is a contraction, thus it suffices to look at

$$f \left[\frac{e^{h(s+r)} - e^{h(s)}}{r} + Ae^{h(s)} \right],$$

but by the definition of $h(s)$ and $h(s+r)$, this equals

$$-f \left[\frac{S(t-s-r)e^h - S(t-s)e^h}{-r} - AS(t-s)e^h \right]$$

which converges to 0 in norm as $r \downarrow 0$.

For the proof of the second statement, let $\{f(s')\}_{s' \leq t} \in C^1([0, t], D)$, then we have for $s \in [0, t]$ that

$$\begin{aligned} & \frac{S^{h[0, t]}(s, s+r)f(s+r) - f(s)}{r} - \left(A^{h(s)}f(s) + \partial f(s) \right) \\ &= \frac{S^{h[0, t]}(s, s+r)f(s) - f(s)}{r} - A^{h(s)}f(s) \\ &+ \frac{S^{h[0, t]}(s, s+r)f(s+r) - S^{h[0, t]}(s, s+r)f(s)}{r} - S^{h[0, t]}(s, s+r)\partial f(s) \\ &+ S^{h[0, t]}(s, s+r)\partial f(s) - \partial f(s). \end{aligned}$$

The first term converges to 0 as shown in the first part of the proof. The second term converges to 0 as $S^{h[0, t]}(s, s+r)$ is contractive for all $r > 0$ and the definition of $\partial f(s)$. The last term converges to 0 as $\{S^{h[0, t]}(r', s')\}_{0 \leq r' \leq s' \leq t}$ is strongly continuous. \square

The next corollary follows directly from the second statement of proposition.

Corollary 4.15. *Let $f \in C^1([0, t], D)$ and $s \in [0, t]$, then*

$$M^f(s) := f(s)(X(s)) - f(0)(X(0)) - \int_0^s A^{h(r)}f(r)(X(r)) + \partial f(r)(X(r))dr$$

is a mean 0 martingale for \mathbb{Q} .

Proposition 4.16. *Suppose $h \in D$, and let ρ_0 be some non-negative measurable function such that $\int \rho_0 d\mathbb{P}_0 = 1$. Define $\mathbb{Q} \in \mathcal{P}(D_E(\mathbb{R}))$ by the change of measure*

$$\frac{d\mathbb{Q}}{d\mathbb{P}}(X) = \rho_0(X(0)) \exp \{h(X(t))\}.$$

Denote with $\gamma(0) \in \mathcal{P}(E)$, the projection of \mathbb{Q} onto the time 0 marginal and denote with $\mathbb{P}^{\gamma(0)} \in \mathcal{P}(D_E(\mathbb{R}^+))$ the Markov measure started from $\gamma(0)$ with transition semigroup $\{S(r)\}_{r \geq 0}$.

Then, we have

$$\begin{aligned} H(\mathbb{Q} | \mathbb{P}) &= H(\mathbb{Q}_0 | \mathbb{P}_0) + H(\mathbb{Q} | \mathbb{P}^{\gamma(0)}) \\ &= H(\mathbb{Q}_0 | \mathbb{P}_0) + \int_0^t \mathcal{L}(\gamma(s), \dot{\gamma}(s))ds, \end{aligned}$$

where for every $s \in \mathbb{R}^+$, $\gamma(s)$ is the law of $X(s)$ under \mathbb{Q} .

Proof. The first equality follows by applying Lemma 4.4.7 in [9] to $H(\mathbb{Q} | \mathbb{P})$ and $H(\mathbb{Q} | \mathbb{P}^{\gamma(0)})$, conditioning with respect to the time 0 marginal, and then subtracting the results.

By construction, we have

$$\frac{d\mathbb{Q}}{d\mathbb{P}^{\gamma(0)}}(X) = e^{h(X(t)) - h(0)(X(0))}$$

so that only the Doob transform part is left. As $H(\mathbb{Q} | \mathbb{P}^{\gamma(0)}) = \int \log \frac{d\mathbb{Q}}{d\mathbb{P}^{\gamma(0)}} d\mathbb{Q}$, we study $h(t)(X(t)) - h(0)(X(0))$. Recall that $\partial h(s) = -Hh(s)$ by Corollary

4.2 and Lemma 4.1:

$$\begin{aligned}
& h(t)(X(t)) - h(0)(X(0)) \\
&= h(t)(X(t)) - h(0)(X(0)) - \int_0^t A^{h(s)}h(s)(X(s)) + \frac{\partial h(s)}{\partial s}(X(s))ds \\
&\quad + \int_0^t A^{h(s)}h(s)(X(s)) + \frac{\partial h(s)}{\partial s}(X(s))ds \\
&= M^h(t) + \int_0^t A^{h(s)}h(s)(X(s)) - Hh(s)(X(s))ds
\end{aligned}$$

where $s \mapsto M^h(s)$ is a mean 0 \mathbb{Q} martingale by Corollary 4.15. Therefore, using Lemma 4.8 in line 3 we see that

$$\begin{aligned}
H(\mathbb{Q} | \mathbb{P}^{\gamma(0)}) &= \int \log \frac{d\mathbb{Q}}{d\mathbb{P}^{\gamma(0)}} d\mathbb{Q} \\
&= \int \int_0^t A^{h(s)}h(s)(X(s)) - Hh(s)(X(s)) ds d\mathbb{Q}(X) \\
&= \int \int_0^t Lh(s)(X(s)) ds d\mathbb{Q}(X).
\end{aligned}$$

By Lemma 2.5, the operator $L : (D, \tau_D) \rightarrow (C(E), \|\cdot\|)$, given by $Lg = A^g g - Hg$ is continuous. As $s \mapsto h(s) = V(t-s)h$ is continuous in (D, τ_D) by Condition 2.3 (d), we see that $s \mapsto Lh(s)$ is norm continuous. Therefore, Fubini's theorem gives us

$$\begin{aligned}
H(\mathbb{Q} | \mathbb{P}^{\gamma(0)}) &= \int_0^t \int Lh(s)(X(s)) d\mathbb{Q}(X) ds \\
&= \int_0^t \langle Lh(s), \gamma(s) \rangle ds \\
&= \int_0^t \mathcal{L}(\gamma(s), \dot{\gamma}(s)) ds.
\end{aligned}$$

□

5 Proof of Theorem 2.8

We proceed with the proof of Theorem 2.8. We start with two crucial compactness results which are necessary for the Nisio semigroup, introduced in Section 5.2, to be well behaved.

5.1 Compactness of the space of paths with bounded Lagrangian cost

We start with proving the compactness of the level sets of \mathcal{L} .

Proposition 5.1. *For each $C \geq 0$, the set*

$$\{(\mu, u) \in \mathcal{P}(E) \times U \mid \mathcal{L}(\mu, u) \leq C\}$$

is compact with respect to the weak topology on $\mathcal{P}(E)$ and the weak topology on U .*

Proof. First of all, as \mathcal{L} is lower semi-continuous $\{(\nu, u) \in \mathcal{P}(E) \times U \mid \mathcal{L}(\nu, u) \leq C\}$ is closed. We show that it is contained in a compact set. Pick the neighbourhood of 0 \mathcal{N} that was given in Condition 2.3 (f), so that $\sup_{f \in \mathcal{N}} \|Hf\| \leq 1$. As $\langle f, u \rangle \leq \mathcal{L}(\mu, u) + \langle Hf, \mu \rangle$, we obtain

$$|\langle f, u \rangle| \leq \mathcal{L}(\nu, u) + \langle Hf, \nu \rangle \vee \langle H(-f), \nu \rangle.$$

As a consequence,

$$\{(\nu, u) \in \mathcal{P}(E) \times U \mid \mathcal{L}(\nu, u) \leq C\} \subseteq \mathcal{P}(E) \times |C + 1|\mathcal{N}^\circ$$

As (D', wk^*) is Hausdorff and a locally convex space, the closure of this set is compact in (D', wk^*) by the Bourbaki-Aloaglu theorem [27, Propositions 32.7 and 32.8], [26, Theorem III.6]. \square

We now state an essential ingredient of the proof of Theorem 2.8.

Proposition 5.2. *For each $M > 0$, and time $T \geq 0$,*

$$\mathcal{K}_M^T := \left\{ \mu \in C_{\mathcal{P}(E)}(\mathbb{R}^+) \mid \mu \in \mathcal{AC}, \int_0^T \mathcal{L}(\mu(s), \dot{\mu}(s)) ds \leq M \right\}$$

is a compact subset of $C_{\mathcal{P}(E)}([0, T])$.

We postpone the lengthy proof of the proposition to Sections 5.4 and 5.5 and focus on proving Theorem 2.8 first.

5.2 The Nisio semigroup

Definition 5.3. The Nisio semigroup \mathbf{V} mapping upper semi-continuous functions on $\mathcal{P}(E)$ to upper semi-continuous functions on $\mathcal{P}(E)$ is defined by

$$\mathbf{V}(t)G(\mu) = \sup_{\nu \in \mathcal{AC}_\mu} \left\{ G(\nu(t)) - \int_0^t \mathcal{L}(\nu(s), \dot{\nu}(s)) ds \right\}.$$

For a function $f \in C(E)$, we denote with $[f]$ the weakly continuous function on $\mathcal{P}(E)$ defined by $[f](\mu) = \langle f, \mu \rangle$. Our goal in this section is to show that $\mathbf{V}(t)[f](\mu) = \langle V(t)f, \mu \rangle$.

Note that as a direct consequence of Proposition 5.2, if G is a bounded continuous function, then the supremum is actually attained by a curve starting at μ in $\mathcal{K}_{3\|G\|}^t$. For example, this is the case if $G = [g]$, for $g \in C(E)$.

We need one small result, that is essential for the analysis. In particular, it is used for the proof of Lemma 5.8.

Lemma 5.4. *For each $\mu \in \mathcal{P}(E)$ and $f \in D$, there exists $\nu \in \mathcal{AC}_\mu$ such that for every $t \geq 0$*

$$\int_0^t \langle f, \dot{\nu}(s) \rangle ds = \int_0^t \langle Hf, \nu(s) \rangle + \mathcal{L}(\nu(s), \dot{\nu}(s)) ds.$$

In particular by taking $f = 0$, we find that there is a path with zero cost. This in turn yields $\mathbf{V}(t)\mathbf{0} = \mathbf{0}$, where $\mathbf{0}$ is the function defined by $\mathbf{0}(\mu) = 0$ for all $\mu \in \mathcal{P}(E)$.

Proof. Let $\nu(s)$ be the path obtained by the time projections of the Markov process started at μ generated by the operator A^f , see for example Theorem 4.2 in Palmowski and Rolski [25]. This gives us a path such that $\dot{\nu}(s) = (A^f)'(\nu(s)) = \rho(\nu(s), f)$.

By Equation (4.5) on page 17, it follows that

$$\langle Hf, \nu(s) \rangle = \langle f, \rho(\nu(s), f) \rangle - \mathcal{L}(\nu(s), \rho(\nu(s), f))$$

for every s , implying that

$$\int_0^t \langle Hf, \nu(s) \rangle ds = \int_0^t (\langle f, \dot{\nu}(s) \rangle - \mathcal{L}(\nu(s), \dot{\nu}(s))) ds.$$

□

The semigroup $\{\mathbf{V}(t)\}_{t \geq 0}$ enjoys good continuity properties.

Lemma 5.5. *For every $t \geq 0$, $\mathbf{V}(t)$ is contractive, i.e. for bounded and upper semi-continuous functions F, G , we have*

$$\|\mathbf{V}(t)F - \mathbf{V}(t)G\| \leq \|F - G\|.$$

The proof of this lemma is straightforward. The next result can be proven using Proposition 5.2 as Lemma 8.16 in [13].

Lemma 5.6. *For every $f \in C(E)$ and $\mu \in \mathcal{P}(E)$, we have that $t \mapsto \mathbf{V}(t)[f](\mu)$ is continuous.*

We proceed with the preparations of Proposition 5.10 where we will prove that $\langle V(t)f, \mu \rangle = \mathbf{V}(t)[f](\mu)$ for $f \in C(E)$ and $\mu \in \mathcal{P}(E)$.

The inequality $\langle V(t)f, \mu \rangle \leq \mathbf{V}(t)[f](\mu)$ is based on the Doob transform method and in particular on Propositions 4.16, 4.14 and 5.2. The other inequality is based on approximation arguments. In the next definition, we introduce the resolvent $\mathbf{R}(\lambda)$ of the Nisio semigroup. Based on Lemma 4.3, we show that $\mathbf{R}(\lambda)[f](\mu) \leq [R(\lambda)f](\mu)$ which by approximation yields $\mathbf{V}(t)[f](\mu) \leq \langle V(t)f, \mu \rangle$.

Definition 5.7. Let G be upper semi-continuous and bounded and let $\lambda > 0$. Define the resolvent $\mathbf{R}(\lambda)$ by

$$\mathbf{R}(\lambda)G(\mu) = \sup_{\nu \in \mathcal{AC}_\mu} \int_0^\infty \frac{1}{\lambda} e^{-\lambda^{-1}s} \left[G(\nu(s)) - \int_0^s \mathcal{L}(\nu(r), \dot{\nu}(r)) dr \right] ds.$$

Lemma 5.8. *For $g \in D$, we have $\mathbf{R}(\lambda)[(1 - \lambda H)g] = [g]$. As a consequence, we have for $f \in C(E)$ and $\mu \in \mathcal{P}(E)$ that*

$$\mathbf{R}(\lambda)[f](\mu) \leq [R(\lambda)f](\mu). \quad (5.1)$$

Proof. The first statement follows along the lines of the proof of Lemma 8.19 in [13]. Summarising, the inequality $\mathbf{R}(\lambda)[(1 - \lambda H)g] \leq [g]$ follows by integration by parts and Youngs inequality:

$$\langle g, u \rangle \leq \langle Hg, \mu \rangle + \mathcal{L}(\mu, u), \quad \mu \in \mathcal{P}(E), u \in D, g \in C(E).$$

The second inequality, $\mathbf{R}(\lambda)[(1 - \lambda H)g] \geq [g]$, follows by integration by parts and Lemma 5.4, which gives us a trajectory for which equality is attained for all times in Youngs inequality.

For the second statement, first note that if $F \geq G$, then $\mathbf{R}(\lambda)F \geq \mathbf{R}(\lambda)G$. Therefore, we obtain by Lemma 4.3 that

$$\mathbf{R}(\lambda)[f](\mu) \leq \mathbf{R}(\lambda)[(1 - \lambda H)R(\lambda)f](\mu) = \langle R(\lambda)f, \mu \rangle.$$

□

The next lemma relies on Lemma 5.6 and follows exactly as Lemma 8.18 in [13].

Lemma 5.9. *For $t \geq 0$, $f \in D$ and $\mu \in \mathcal{P}(E)$, we have*

$$\lim_{n \rightarrow \infty} (n\mathbf{R}(n))^{\lfloor nt \rfloor} [f](\mu) = \mathbf{V}(t)[f](\mu).$$

We are now able to prove the important result that identifies the Nisio semigroup with $\{V(t)\}_{t \geq 0}$.

Proposition 5.10. *For $t \geq 0$, $f \in C(E)$ and $\mu \in \mathcal{P}(E)$, we have*

$$\mathbf{V}(t)[f](\mu) = \langle V(t)f, \mu \rangle.$$

Proof. By repeatedly using Equation (5.1), we obtain

$$\mathbf{R}(n^{-1})^{\lfloor nt \rfloor} [f](\mu) \leq \langle R(n^{-1})^{\lfloor nt \rfloor} f, \mu \rangle,$$

which implies by Lemmas 4.4 and 5.9 that

$$\mathbf{V}(t)[f](\mu) \leq \langle V(t)f, \mu \rangle. \quad (5.2)$$

For the second inequality, we first pick $f \in D$. Let $\mathbb{P} \in \mathcal{P}(D_E(\mathbb{R}^+))$ be the Markov measure started from $\mu \in \mathcal{P}(E)$ with transition semigroup $\{S(t)\}_{t \geq 0}$.

By Proposition 3.7, we find a measure $\mathbb{Q}^* \in \mathcal{P}(D_E(\mathbb{R}^+))$ and a sequence of measures \mathbb{Q}^n such that

$$\begin{aligned} \langle V(t)f, \mu \rangle &= \sup_{\nu \in \mathcal{P}(E)} \{ \langle f, \nu \rangle - I_t(\nu | \mu) \} \\ &= \langle f, \phi_t(\mathbb{Q}^*) \rangle - H(\mathbb{Q}^* | \mathbb{P}) \\ &= \lim_{n \rightarrow \infty} \langle f, \phi_t(\mathbb{Q}_n) \rangle - H(\mathbb{Q}^n | \mathbb{P}) \end{aligned}$$

Let $\gamma^n \in \mathcal{AC}$ be the measure valued trajectories obtained from the laws of \mathbb{Q}^n . Proposition 4.16 yields

$$\begin{aligned} \langle V(t)f, \mu \rangle &= \lim_{n \rightarrow \infty} \langle f, \phi_t(\mathbb{Q}_n) \rangle - H(\mathbb{Q}_0^n | \mathbb{P}_0) - \int_0^t \mathcal{L}(\gamma^n(s), \dot{\gamma}^n(s)) ds \\ &= \lim_{n \rightarrow \infty} \langle f, \phi_t(\mathbb{Q}_n) \rangle - \int_0^t \mathcal{L}(\gamma^n(s), \dot{\gamma}^n(s)) ds \end{aligned}$$

as $H(\mathbb{Q}_0^n | \mathbb{P}_0) \rightarrow 0$. For all $n \in \mathbb{N}$ we have $\langle f, \gamma_n(t) \rangle = \alpha_0$. Also, we have for all $n \in \mathbb{N}$ that $\int_0^t \mathcal{L}(\gamma_n(s), \dot{\gamma}_n(s)) ds \leq H(\mathbb{Q}^n | \mathbb{P}) \leq H(\mathbb{Q}^* | \mathbb{P})$, which implies by Proposition 5.1 that we can find a subsequence $k \mapsto n(k)$ in \mathbb{N} such that $\gamma_{n(k)}$

converges in $C_{\mathcal{P}(E)}(\mathbb{R}^+)$ to a trajectory $\gamma \in \mathcal{AC}$ that starts at μ . Therefore, we obtain by using the lower semi-continuity of \mathcal{L} in line two that

$$\begin{aligned} \langle V(t)f, \mu \rangle &= \lim_{n \rightarrow \infty} \left\{ \langle f, \gamma_n(t) \rangle - H(\mathbb{Q}_0^* | \mu) - \int_0^t \mathcal{L}(\gamma_n(s), \dot{\gamma}_n(s)) ds \right\} \\ &\leq \langle f, \gamma(t) \rangle - \int_0^t \mathcal{L}(\gamma(s), \dot{\gamma}(s)) ds \\ &\leq \mathbf{V}(t)[f](\mu). \end{aligned}$$

This inequality, together with (5.2) yields $\langle V(t)f, \mu \rangle = \mathbf{V}(t)[f](\mu)$ for $f \in D$. The result for $f \in C(E)$ follows by the continuity of $f \mapsto V(t)f$ and the continuity of $f \mapsto \mathbf{V}(t)f$ given by Lemma 5.5. \square

5.3 The Lagrangian form of the rate function

In this section, we show that I_t can be re-expressed using the Nisio semigroup.

Lemma 5.11. *Under the Condition 2.3, it holds that*

$$I_t(\mu_1 | \mu_0) = \inf_{\substack{\nu \in \mathcal{AC}_{\mu_0} \\ \nu(t) = \mu_1}} \int_0^t \mathcal{L}(\nu(s), \dot{\nu}(s)) ds.$$

The proof is a classical proof using convex duality.

Proof. For a fixed measure $\mu_0 \in \mathcal{P}(E)$, consider the function $\mathbb{L}_{\mu_0} : \mathcal{P}(E) \rightarrow \mathbb{R}^+$ defined by

$$\mathbb{L}_{\mu_0}(\mu_1) := \inf_{\substack{\nu \in \mathcal{AC}_{\mu_0} \\ \nu(t) = \mu_1}} \int_0^t \mathcal{L}(\nu(s), \dot{\nu}(s)) ds$$

Our goal is to prove that $I_t(\mu_1 | \mu_0) = \mathbb{L}_{\mu_0}(\mu_1)$ by showing that both are the Fenchel-Legendre transform of $\langle V(t)g, \mu_1 \rangle$. First, we will prove that \mathbb{L}_{μ_0} is convex and has compact level sets. This last result implies the lower semi-continuity.

Step 1. The convexity of \mathbb{L}_{μ_0} follows directly from the convexity of \mathcal{L} and the fact that \mathcal{AC} is convex. So we are left to prove compactness of the level sets. Pick a sequence μ^n in the set $\{\mu | \mathbb{L}_{\mu_0}(\mu) \leq c\}$. We know by definition of \mathbb{L}_{μ_0} and Proposition 5.2 that there are $\nu^n \in \mathcal{K}_{c, \{\mu_0\}}^t$ such that $\nu^n(0) = \mu_0$, $\nu^n(t) = \mu^n$ and

$$\int_0^t \mathcal{L}(\nu^n(s), \dot{\nu}^n(s)) ds \leq c.$$

Again by Proposition 5.2, we obtain that the sequence ν^n has a converging subsequence ν^{n_k} with limit ν^* such that

$$\int_0^t \mathcal{L}(\nu^*(s), \dot{\nu}^*(s)) ds \leq c.$$

Denote with $\mu^* := \nu^*(t)$, then we know that $\nu^{n_k}(t) \rightarrow \mu^*$ and $\mathbb{L}_{\mu_0}(\mu^*) \leq c$, which implies that $\mathbb{L}_{\mu_0}(\cdot)$ has compact level sets and is lower semi-continuous.

Step 2. Now that we know that \mathbb{L}_{μ_0} is convex and lower semi-continuous, we are able to prove that $\mathbb{L}_{\mu_0}(\cdot) = I_t(\cdot | \mu_0)$.

$\mathbb{L}_{\mu_0}(\cdot)$ is lower semi-continuous on $\mathcal{P}(E)$ with respect to the weak topology, so extending its domain of definition to $\mathcal{M}(E)$ by setting it equal to ∞ outside $\mathcal{P}(E)$ does not change the fact that it is lower semi-continuous.

Because the dual of $(\mathcal{M}(E), \text{weak})$ is $C(E)$ by the Riesz representation theorem and [4, Theorem V.1.3], we obtain by Lemma 4.5.8 in Dembo and Zeitouni [8] that the Fenchel-Legendre transform of

$$\begin{aligned} \sup_{\mu_1} \{ \langle g, \mu_1 \rangle - \mathbb{L}_{\mu_0}(\mu_1) \} &= \sup_{\nu \in \mathcal{AC}_{\mu_0}} \left\{ \langle g, \nu(t) \rangle - \int_0^t \mathcal{L}(\nu(s), \dot{\nu}(s)) ds \right\} \\ &= \mathbf{V}(t)[g](\mu_0) \end{aligned}$$

satisfies $\mathbb{L}_{\mu_0}(\mu_1) = \sup_{g \in C(E)} \{ \langle g, \mu_1 \rangle - \mathbf{V}(t)[g](\mu_0) \}$. Therefore, by Proposition 5.10, we see

$$\mathbb{L}_{\mu_0}(\mu_1) = \sup_{g \in C_0(E)} \{ \langle g, \mu_1 \rangle - \langle V(t)g, \mu_0 \rangle \}. \quad (5.3)$$

On the other hand, by Theorem 2.1,

$$I_t(\mu_1 | \mu_0) = \sup_{g \in C_0(E)} \{ \langle g, \mu_1 \rangle - \langle V(t)g, \mu_0 \rangle \}. \quad (5.4)$$

The combination of Equations (5.3) and (5.4), i.e. both are the Legendre-Fenchel transform of $\langle V(t)g, \mu_0 \rangle$, yields that

$$I_t(\mu_1 | \mu_0) = \mathbb{L}_{\mu_0}(\mu_1) = \inf_{\substack{\nu \in \mathcal{AC}_{\mu_0} \\ \nu(t) = \mu_1}} \int_0^t \mathcal{L}(\nu(s), \dot{\nu}(s)) ds.$$

□

We proceed with the final lemma before the proof of Theorem 2.8.

Lemma 5.12. *The function $J : C_{\mathcal{P}(E)}(\mathbb{R}^+) \rightarrow \mathbb{R}$, given by*

$$J(\mu) = \begin{cases} H(\mu(0) | \mathbb{P}_0) + \int_0^\infty \mathcal{L}(\mu(s), \dot{\mu}(s)) ds & \text{if } \mu \in \mathcal{AC}, \mu(0) < \mathbb{P}_0 \\ \infty & \text{otherwise,} \end{cases}$$

has compact level sets in $C_{\mathcal{P}(E)}(\mathbb{R}^+)$.

Proof. Clearly, $\{J \leq M\} \subseteq \bigcap_T \mathcal{K}_M^T$. So, pick a sequence $\mu^n \in \{J \leq M\}$. For $n = 1$, we can construct a converging subsequence μ^{n_k} in \mathcal{K}_M^1 seen as a subset of $C_{\mathcal{P}(E)}([0, 1])$. From this subsequence, we can extract yet another subsequence that has the same property on $[0, 2]$. By a diagonal argument, this yields a converging subsequence in $C_{\mathcal{P}(E)}(\mathbb{R}^+)$. By the lower semi-continuity of $H(\cdot | \mathbb{P}_0)$ and \mathcal{L} this yields that the limit is in $\{J \leq M\}$. □

Proof of Theorem 2.8. By using the contraction principle from the space

$$C_{\mathcal{P}(E)}(\mathbb{R}^+) \rightarrow \prod_{\mathbb{R}^+} \mathcal{P}(E)$$

using the identity map, we find that the rate function in Theorem 2.1 coincides with the rate function which would have been found via the Dawson-Gärtner theorem [8, Theorem 4.6.1] for the large deviation problem on $\prod_{\mathbb{R}^+} \mathcal{P}(E)$.

In this context, we can apply Lemma 4.6.5 [8] to find that if we have a good rate function J on $\prod_{\mathbb{R}^+} \mathcal{P}(E)$ that satisfies

$$I[0, t_1, \dots, t_k](\mu(0), \mu(t_1), \dots, \mu(t_k)) = \inf \{J(\nu) \mid \nu(0) = \mu(0), \nu(t_i) = \mu(t_i)\}, \quad (5.5)$$

then it holds that $I = J$. The candidate

$$J(\mu) = \begin{cases} H(\mu(0) \mid \mathbb{P}_0) + \left\{ \int_0^\infty \mathcal{L}(\mu(s), \dot{\mu}(s)) ds \right\} & \text{if } \mu \in \mathcal{AC}_{\mu_0}, \mu(0) << \mathbb{P}_0 \\ \infty & \text{otherwise,} \end{cases}$$

clearly satisfies Equation (5.5). By Lemma 5.12, we know that J is a good rate function on $C_{\mathcal{P}(E)}(\mathbb{R}^+)$ and therefore also on $\prod_{\mathbb{R}^+} \mathcal{P}(E)$. \square

5.4 Preparations for the proof of Proposition 5.2

We say that a topological space is Souslin if it is the continuous image of a complete separable metric space. For the proof of Proposition 5.2, we will need the generalisation of one of the implications of the Prohorov theorem.

Theorem 5.13 (Prohorov). *Let \mathcal{K} be a subset of the Borel measures on a completely regular Souslin space \mathcal{S} that is uniformly bounded with respect to the total variation norm. If \mathcal{K} is a tight family of measures, then \mathcal{K} has a compact and sequentially compact closure with respect to the weak topology on $\mathcal{P}(\mathcal{S})$.*

The Prohorov theorem is given in [1, Theorem 8.6.7] and its specialisation to completely regular Souslin spaces follows from [1, Corollary 6.7.8 and Theorem 7.4.3]

Remark 5.14. The other implication of the ordinary Prohorov theorem does not necessarily hold in this generality [1, Proposition 8.10.19].

We will use the Prohorov theorem for measures on the product space $(\mathcal{P}(E) \times U \times [0, T])$, where the first two spaces are equipped with the weak* topology, and the last space with its standard topology.

Lemma 5.15. *The space $(\mathcal{P}(E) \times U \times [0, T])$ is completely regular and Souslin.*

Proof. We start with proving that $(\mathcal{P}(E) \times U \times [0, T])$ is completely regular. By Lemma [19, 15.2.(3)] (D', wk^*) is completely regular, therefore, the subspace (U, wk^*) is completely regular. This yields the result as taking products preserves complete regularity.

By Condition 2.3 (a) and Lemma 8.6, we obtain that (U, wk^*) is Souslin. Clearly, $(\mathcal{P}(E), wk)$ and $[0, T]$ are Souslin, so that the product space $(\mathcal{P}(E) \times U \times [0, T])$ is Souslin by Lemma 6.6.5 in Bogachev [1]. \square

Suppose that we have a weakly converging net of measures on $(\mathcal{P}(E) \times U \times [0, T])$. By definition, integrals of continuous and bounded functions with respect to this net of measures converges in \mathbb{R} . The next lemmas are aimed to extend this property to continuous functions, that are unbounded, but linear on U .

Definition 5.16. For the neighbourhood \mathcal{N} , we define the Minkowski functional on U

$$\|u\|_{\mathcal{N}} := \inf \{c \geq 0 \mid u \in c\mathcal{N}^\circ\}.$$

We have the following elementary results.

Lemma 5.17. $\|\cdot\|_{\mathcal{N}}$ is a norm on U , $\{u \mid \|u\|_{\mathcal{N}} \leq 1\} = \mathcal{N}^\circ$. Furthermore, for $u \in U$, we have

$$\sup_{f \in c\mathcal{N}} \frac{\langle f, u \rangle}{\|u\|_{\mathcal{N}}} = c.$$

We use this lemma to find functions ϕ of the type given in the following lemma, which is an analogue of the de la Vallée-Poussin lemma [1, Theorem 4.5.9].

Lemma 5.18. For a net of measures π^α bounded in total variation norm, that weakly converges to a measure π , and a measurable function f , suppose that there exists a non-negative non-decreasing function $\phi : \mathbb{R}^+ \rightarrow \mathbb{R}^+$ which satisfies

$$\lim_{r \rightarrow \infty} \frac{\phi(r)}{r} = \infty,$$

for which it holds that $\sup_\alpha \int \phi(|f|) d\pi^\alpha \leq M < \infty$, then it holds that

$$\sup_\alpha \int |f| d\pi^\alpha < \infty.$$

Also, we obtain that uniformly in α

$$\lim_{C \rightarrow \infty} \left| \int f d\pi^\alpha - \int \Upsilon_C(f) d\pi^\alpha \right| = 0, \quad (5.6)$$

where $\Upsilon_C(f) = (f \vee -C) \wedge C$.

Proof. Fix $\varepsilon > 0$ and pick $C(\varepsilon)$ big enough such that for $r \geq C(\varepsilon)$ we have $\frac{\phi(r)}{r} \geq \frac{M}{\varepsilon}$. Then, we obtain that

$$\sup_\alpha \int_{|f| \geq C} |f| d\pi^\alpha \leq \sup_\alpha \frac{\varepsilon}{M} \int_{|f| \geq C} \phi(|f|) d\pi^\alpha \leq \frac{\varepsilon}{M} M \leq \varepsilon.$$

As a consequence, we see

$$\sup_\alpha \int |f| d\pi^\alpha \leq C(\varepsilon) \sup_\alpha \|\pi^\alpha\|_{TV} + \varepsilon < \infty.$$

The second statement follows by the observation that

$$\sup_\alpha \int |f - \Upsilon_C(f)| d\pi^\alpha \leq \sup_\alpha \int_{|f| \geq C} |f| d\pi^\alpha.$$

□

Lemma 5.19. Under Condition 2.3 (f) that states that for every $c \geq 0$: $\Gamma(c) := \sup_{f \in c\mathcal{N}} \|Hf\| < \infty$, there exists a non-decreasing function $\phi : \mathbb{R}^+ \rightarrow \mathbb{R}^+$, such that $\lim_{r \rightarrow \infty} \frac{\phi(r)}{r} = \infty$ and such that $\phi(|\langle f, u \rangle|) \leq \phi(\|u\|_{\mathcal{N}}) \leq \mathcal{L}(\mu, u)$ for every $f \in \mathcal{N}$, $u \in U$ and $\mu \in \mathcal{P}(E)$.

The proof of this lemma is inspired by the proof of Lemma 10.21 in Feng and Kurtz [13].

Proof. For $u \neq 0$ in U , Lemma 5.17 yields

$$\frac{\mathcal{L}(\mu, u)}{\|u\|_{\mathcal{N}}} \geq \sup_{f \in c\mathcal{N}} \left\{ \frac{\langle f, u \rangle}{\|u\|_{\mathcal{N}}} - \frac{\langle Hf, \mu \rangle}{\|u\|_{\mathcal{N}}} \right\} \geq c - \frac{\Gamma(c)(c)}{\|u\|_{\mathcal{N}}}$$

for every $c > 0$. This directly yields for every $c > 0$

$$\lim_{r \rightarrow \infty} \inf_{\mu \in \mathcal{P}(E)} \inf_{u : \|u\|_{\mathcal{N}} \geq r} \frac{\mathcal{L}(\mu, u)}{\|u\|_{\mathcal{N}}} \geq \lim_{r \rightarrow \infty} \inf_{\mu \in \mathcal{P}(E)} \inf_{u : \|u\|_{\mathcal{N}} \geq r} c - \frac{\Gamma(c)(c)}{\|u\|_{\mathcal{N}}} = c,$$

which implies

$$\lim_{r \rightarrow \infty} \inf_{\mu \in \mathcal{P}(E)} \inf_{u : \|u\|_{\mathcal{N}} \geq r} \frac{\mathcal{L}(\mu, u)}{\|u\|_{\mathcal{N}}} = \infty.$$

Consequently, the function

$$\phi(r) = r \inf_{\mu \in \mathcal{P}(E)} \inf_{u : \|u\|_{\mathcal{N}} \geq r} \frac{\mathcal{L}(\mu, u)}{\|u\|_{\mathcal{N}}},$$

satisfies the claims in the lemma. \square

5.5 Proof of Proposition 5.2

We now have the tools for the proof of Proposition 5.2. Essentially, the proof follows the approach as in Feng and Kurtz [13, Proposition 8.13]. We give it for clarity and completeness as there are some notable differences. First of all, we work with absolutely continuous paths, instead of paths that satisfy a relaxed control equation. Second, the possible ‘speeds’ that we allow are elements of the completely regular Souslin subset U of a locally convex space instead of a metric space.

Proof of Proposition 5.2. Pick a sequence $\mu^n \in \mathcal{K}_M^T$. As $\mathcal{P}(E)$ is compact, we assume that $\mu^n(0) \rightarrow \mu_0$. Define the occupation measures π^n on $\mathcal{P}(E) \times U \times [0, T] \subseteq \mathcal{P}(E) \times U \times [0, T]$ by

$$\pi^n(C \times [0, t]) = \int_0^t \mathbb{1}_C(\mu^n(s), \dot{\mu}^n(s)) ds.$$

Proposition 5.1 tells us that π^n is tight in $\mathcal{P}(\mathcal{P}(E) \times U \times [0, T])$ by considering the following calculation:

$$\begin{aligned} & C \pi^n \{(\mu, u, t) \in \mathcal{P}(E) \times U \times [0, T] \mid \mathcal{L}(\mu, u) \leq C\}^c \\ & \leq \int_0^T \mathcal{L}(\mu, u) \pi^n(d\mu \times du \times ds) \\ & \leq M. \end{aligned}$$

In other words

$$\pi^n \{(\mu, u, t) \in \mathcal{P}(E) \times U \times [0, T] \mid \mathcal{L}(\mu, u) \leq C\}^c \leq \frac{M}{C}, \quad (5.7)$$

and because C is arbitrary, we can choose it big enough such that this probability is smaller than any $\varepsilon > 0$ uniformly in n . This implies by Theorem 5.13 that π^n

contains a weakly converging subsequence. Therefore, we assume without loss of generality that, there exists $\pi \in \mathcal{P}(\hat{K} \times U \times [0, T])$ such that $\pi^n \rightarrow \pi$ weakly. We now show that π gives us a new path $s \mapsto \mu(s)$ in \mathcal{K}_M^T . Recall that for $c \geq 0$ $\Upsilon_c(g) = (g \wedge c) \vee -c$. So for a fixed $f \in D$, $u \mapsto \Upsilon_c(\langle f, u \rangle)$ is a bounded and continuous function. For an arbitrary $t \leq T$, the set $\pi(\mathcal{P}(E) \times U \times \{t\})$ is a set of measure 0, so the function $(u, s) \mapsto \mathbb{1}_{\{s \leq t\}} \Upsilon_c(\langle f, u \rangle)$ is a bounded Borel measurable functions that is continuous π almost everywhere. Hence, by the weak convergence of π^n to π and Corollary 8.4.2 in Bogachev [1], we obtain for every $c \geq 0$ that

$$\int_{\{s \leq t\}} \Upsilon_c(\langle f, u \rangle) \pi^n(d\mu \times du \times ds) \rightarrow \int_{\{s \leq t\}} \Upsilon_c(\langle f, u \rangle) \pi(d\mu \times du \times ds).$$

By the Portmanteau theorem and the lower semi-continuity of \mathcal{L} , we obtain that

$$\int \mathcal{L}(\mu, u) \pi(d\mu \times du \times ds) \leq \liminf_n \int \mathcal{L}(\mu, u) \pi^n(d\mu \times du \times ds) \leq M.$$

As $\phi(|\langle f, u \rangle|) \leq \mathcal{L}(\mu, u)$ by Lemma 5.19, and the fact that ϕ satisfies the conditions of Lemma 5.18, we use the result in (5.6) to obtain that

$$\sup_n \left| \int_{\{s \leq t\}} \langle f, u \rangle \pi^n(d\mu \times du \times ds) - \int_{\{s \leq t\}} \Upsilon_c(\langle f, u \rangle) \pi^n(d\mu \times du \times ds) \right| \rightarrow 0,$$

as $c \rightarrow \infty$. This also follows for the limiting measure π :

$$\left| \int_{\{s \leq t\}} \langle f, u \rangle \pi(d\mu \times du \times ds) - \int_{\{s \leq t\}} \Upsilon_c(\langle f, u \rangle) \pi(d\mu \times du \times ds) \right| \rightarrow 0.$$

Thus, by first sending c and then n to infinity, we get

$$\begin{aligned} & \left| \int_{\{s \leq t\}} \langle f, u \rangle \pi^n(d\mu \times du \times ds) - \int_{\{s \leq t\}} \langle f, u \rangle \pi(d\mu \times du \times ds) \right| \\ & \leq \left| \int_{\{s \leq t\}} \langle f, u \rangle \pi^n(d\mu \times du \times ds) - \int_{\{s \leq t\}} \Upsilon_c(\langle f, u \rangle) \pi^n(d\mu \times du \times ds) \right| \\ & \quad + \left| \int_{\{s \leq t\}} \Upsilon_c(\langle f, u \rangle) \pi^n(d\mu \times du \times ds) - \int_{\{s \leq t\}} \Upsilon_c(\langle f, u \rangle) \pi(d\mu \times du \times ds) \right| \\ & \quad + \left| \int_{\{s \leq t\}} \Upsilon_c(\langle f, u \rangle) \pi(d\mu \times du \times ds) - \int_{\{s \leq t\}} \langle f, u \rangle \pi(d\mu \times du \times ds) \right| \\ & \rightarrow 0. \end{aligned} \tag{5.8}$$

Fix some $0 \leq t \leq T$ and pick a sequence $0 \leq t_n \leq T$ that converges to t . Because $\mu^n(t_n)$ is a sequence in the compact set $\mathcal{P}(E)$ it has a converging subsequence with limit ν . By Lemmas 5.18, 5.19, and the Dominated convergence theorem, we have

$$\lim_{n \rightarrow \infty} \int \mathbb{1}_{\{s \text{ between } t_n \text{ and } t\}} |\langle f, u \rangle| \pi^n(d\mu \times du \times ds) \rightarrow 0,$$

which implies, using Equation (5.8), that

$$\begin{aligned}
& \langle f, \nu \rangle - \langle f, \mu_0 \rangle \\
&= \lim_n \langle f, \mu^n(t_n) \rangle - \langle f, \mu^n(0) \rangle \\
&= \lim_n \int \mathbb{1}\{s \leq t\} \langle f, u \rangle \pi^n(d\mu \times du \times ds) \\
&\quad - \int \mathbb{1}\{s \text{ between } t_n \text{ and } t\} \langle f, u \rangle \pi^n(d\mu \times du \times ds) \\
&= \int \mathbb{1}\{s \leq t\} \langle f, u \rangle \pi(d\mu \times du \times ds).
\end{aligned}$$

As D is dense in $C(E)$, this uniquely determines ν , and for every sequence $s_n \rightarrow t$, one gets $\mu^n(s_n) \rightarrow \nu$ weakly. Therefore, we will denote $\mu(t) := \nu$. This way, we can construct $\mu(t)$ for a countable dense subset J of $[0, T]$ and $\mu(t)$ is continuous on J . As a consequence, $\mu(t)$ extends continuously to $[0, t]$ and satisfies

$$\langle f, \mu(t) \rangle - \langle f, \mu_0 \rangle = \int \mathbb{1}_{\{s \leq t\}} \langle f, u \rangle \pi(d\mu \times du \times ds)$$

for every $f \in D$. This implies that for any sequence $s_n \rightarrow t$, we have $\mu(s_n) \rightarrow \mu(t)$, which yields that $\{\mu^n(t)\}_{0 \leq t \leq T}$ converges to $\{\mu(t)\}_{0 \leq t \leq T}$ in $C_{\mathcal{P}(E)}([0, T])$. We proceed with extracting the speed of the trajectory $s \mapsto \mu(s)$ from the measure π . Let $\hat{\pi}$ be the measure π restricted to $U \times [0, T]$. By Corollary 10.4.6 in Bogachev [1], we can write $\hat{\pi}(du \times ds)$ as $\lambda_s(du)ds$.

For Lebesgue almost every s , we know that $\int |\langle f, u \rangle| \lambda_s(du) < \infty$, so we can define the Gelfand integral $\bar{u}(s) = \int u \lambda_s(du)$, see Theorem 8.4. We show that $\bar{u}(s) = \dot{\mu}(s)$. First, by the measurability of $s \mapsto \lambda_s$, also $s \mapsto \bar{u}$ is measurable. Second, by Jensen's inequality in the first line, and the lower semi-continuity of \mathcal{L} in the third,

$$\begin{aligned}
\int_0^T |\langle f, \bar{u}(s) \rangle| ds &\leq \int |\langle f, u \rangle| \pi(d\mu \times du \times ds) \\
&\leq T(\|Hf\| \vee \|H(-f)\|) + \int \mathcal{L}(\mu, u) \pi(d\mu \times du \times ds) \\
&\leq T(\|Hf\| \vee \|H(-f)\|) + \liminf_n \int \mathcal{L}(\mu, u) \pi^n(d\mu \times du \times ds) \\
&\leq T(\|Hf\| \vee \|H(-f)\|) + M.
\end{aligned}$$

Last,

$$\begin{aligned}
\langle f, \mu(t) \rangle - \langle f, \mu(0) \rangle &= \int \mathbb{1}_{\{s \leq t\}} \langle f, u \rangle \pi(d\mu \times du \times ds) \\
&= \int \mathbb{1}_{\{s \leq t\}} \langle f, u \rangle \hat{\pi}(d\mu \times du \times ds) \\
&= \int_0^t \int \langle f, u \rangle \lambda_s(du) ds \\
&= \int_0^t \langle f, \bar{u}(s) \rangle ds.
\end{aligned}$$

This means that $\mu \in \mathcal{AC}^T$ and $\dot{\mu} = \bar{u}$.

We still need to show that $\mu \in \mathcal{K}_M^T$. By the construction of the path $s \mapsto \mu(s)$, it is clear that we have $\pi(d\mu \times du \times ds) = \mathbb{1}_{\{s \leq T\}} \delta_{\{\mu(s)\}}(d\mu) \lambda_s(du) ds$. This shows, using the convexity of \mathcal{L} in the second line, and lower semi-continuity of \mathcal{L} in the third line, that

$$\begin{aligned} \int_0^T \mathcal{L}(\mu(s), \dot{\mu}(s)) ds &= \int \mathcal{L}(\mu, u) \mathbb{1}_{\{s \leq T\}} \delta_{\mu(s)}(d\mu) \delta_{\bar{u}(s)}(du) ds \\ &\leq \int \mathcal{L}(\mu, u) \mathbb{1}_{\{s \leq T\}} \delta_{\mu(s)}(d\mu) \lambda_s(du) ds \\ &\leq \liminf_n \int_0^T \mathcal{L}(\mu^n(s), \dot{\mu}^n(s)) ds \\ &\leq M. \end{aligned}$$

So indeed \mathcal{K}_M^T is compact in $C_{\mathcal{P}(E)}(\mathbb{R}^+)$. \square

6 Examples

We give a number of examples on which Theorem 2.8 can be applied. First of all, we begin with a Markov jump process on a compact metric space. After that, interacting particle systems [23] are considered. In that case, we also prove a representation theorem for D' . Finally, we consider diffusion processes.

6.1 Markov pure jump process

On a compact metric space (E, d) , we have a Markov process $X(t)$ with associated semigroup $S(t) : C(E) \rightarrow C(E)$ generated by the bounded generator

$$Af(a) = \int r(a, db) [f(b) - f(a)],$$

where for every a $r(a, \cdot)$ is some non-negative measure, which is weakly continuous in a , satisfying $\|r\|_\infty = \sup_a r(a, E) < \infty$. We work with the space $(D, \tau_D) = (C(E), \|\cdot\|)$. In this case, the generators A^g and operator H are given by

$$\begin{aligned} A^g f(a) &= \int r(a, db) e^{g(b) - g(a)} [f(b) - f(a)], \\ Hf(a) &= \int r(a, db) [e^{f(b) - f(a)} - 1]. \end{aligned}$$

Lemma 6.1. *Conditions 2.2 and 2.3 are satisfied.*

Proof. Take $D = C(E)$, which clearly satisfies Conditions 2.2 (a) and (b').

Conditions 2.3 (a)-(c), (e) are clear. For (d), we only need to prove that $t \mapsto V(t)f$ is continuous for every $f \in C(E)$. So take a sequence $t_n \in \mathbb{R}^+$ converging to $t \in \mathbb{R}^+$. Then $S(t_n)e^f \rightarrow S(t)e^f$ by the strong continuity of $\{S(t)\}_{t \geq 0}$. As f is bounded, the functions $S(t_n)e^f$ satisfy $e^{-\|f\|} \leq S(t_n)e^f(x) \leq e^{\|f\|}$ for all $x \in E$. On $[e^{-\|f\|}, e^{\|f\|}]$ the logarithm is uniformly continuous, which implies that $\|V(t_n)f - V(t)f\| \rightarrow 0$.

Finally, (f) is satisfied by taking $\mathcal{N} = \{g \in C(E) \mid \|g\| \leq \frac{1}{2} \log(\|r\|^{-1} + 1)\}$. \square

6.2 Interacting particle systems

Let W be a compact metric space and let S be a countable set. Define $(E = W^S, d)$, the product space with d a metric that is compatible with the topology, on which we would like to define a Markov process $\{\eta(t)\}_{t \geq 0}$. Examples are the exclusion process, the contact process, etcetera. We follow the construction in Liggett [23].

For Λ a finite subset of S and $\zeta \in W^\Lambda$ let $c_\Lambda(\eta, d\zeta)$ be the rate at which the system makes a transformation from configuration η to η^ζ which is defined by

$$\eta_x^\zeta = \begin{cases} \eta_x & \text{if } x \notin \Lambda, \\ \zeta_x & \text{if } x \in \Lambda. \end{cases}$$

Put $c_\Lambda = \sup\{c_\Lambda(\eta, W^\Lambda) \mid \eta \in E\}$, the maximal total variation of $c_\Lambda(\eta, \cdot)$. We assume that $c_\Lambda(\eta, d\zeta)$ is weakly continuous in the first variable. We define for finite $\Lambda \subseteq S$ and $u \in S$:

$$c_\Lambda(u) = \sup \{ \|c_\Lambda(\eta, d\zeta) - c_\Lambda(\hat{\eta}, d\zeta)\|_{TV} \mid \eta_y = \hat{\eta}_y \text{ for } y \neq u \},$$

where $\|\cdot\|_{TV}$ refers to the total variation norm. This measures the amount that $\eta \mapsto c_\Lambda(\eta, \cdot)$ depends on the coordinate η_u . Furthermore, let $\gamma(x, u) = \sum_{\Lambda \ni x} c_\Lambda(u)$ for $u \neq x$ and $\gamma(x, x) = 0$ for all x . For $f \in C(E)$, define

$$\Delta_f(x) = \sup \{ |f(\eta) - f(\zeta)| \mid \text{for } y \neq x : \eta_y = \zeta_y \}$$

the variation of f at $x \in S$. For a function in $C(E)$ let $\mathcal{D}(f) := \{x \in S \mid \Delta_f(x) > 0\}$ be the dependence set of f and define the space of local functions by

$$\{f \in D \mid |\mathcal{D}(f)| < \infty\}.$$

and the space of test functions by

$$D = \left\{ f \in C_b(E) \mid \|f\| := \sum_{x \in S} \Delta_f(x) < \infty \right\}, \quad (6.1)$$

which is the closure of the space of local functions with respect to the $\|\cdot\|$ semi-norm.

For functions $f \in D$, define the formal generator A to be

$$Af(\eta) = \sum_{\Lambda} \int c_\Lambda(\eta, d\zeta) [f(\eta^\zeta) - f(\eta)]. \quad (6.2)$$

Theorem I.3.9 in Liggett [23] shows that the closure of A generates a Feller semigroup $\{S(t)\}_{t \geq 0}$. Using this semigroup a Markov process $(\eta(t))_{t \geq 0}$ is constructed such that $S(t)f(\eta) = \mathbb{E}[f(\eta(t)) \mid \eta(0) = \eta]$.

Theorem 6.2 (Liggett I.3.9). *Assume that*

$$\sup_x \sum_{\Lambda \ni x} c_\Lambda < \infty, \quad (6.3)$$

and

$$M := \sup_{x \in S} \sum_{\Lambda \ni x} \sum_{u \neq x} c_\Lambda(u) = \sup_{x \in S} \sum_u \gamma(x, u) < \infty. \quad (6.4)$$

Finally, define the quantity

$$\varepsilon = \inf_{u \in S} \inf_{\substack{\eta_1 = \eta_2 \\ \text{off } u \\ \eta_1(u) \neq \eta_2(u)}} \sum_{\Lambda \ni u} [c_\Lambda(\eta_1, \{\zeta \mid \zeta(u) = \eta_2(u)\}) + c_\Lambda(\eta_2, \{\zeta \mid \zeta(u) = \eta_1(u)\})].$$

Then, we have the following:

- (a) The closure of \overline{A} of A generates a strongly continuous positive contraction semigroup $S(t)$.
- (b) D is a core for \overline{A} .
- (c) If $f \in D$, then $S(t)f \in D$ for all $t \geq 0$ and

$$\|S(t)f\| \leq e^{t(M-\varepsilon)} \|f\|.$$

To make the notation a bit easier, we do not distinguish between \overline{A} and A . A calculation gives the expressions for $A^g f$ and Hf for $f, g \in D$.

$$\begin{aligned} A^g f(\eta) &= \sum_{\Lambda} \int c_\Lambda(\eta, d\zeta) e^{g(\eta^\zeta) - g(\eta)} [f(\eta^\zeta) - f(\eta)] \\ Hf(\eta) &= \sum_{\Lambda} \int c_\Lambda(\eta, d\zeta) [e^{g(\eta^\zeta) - g(\eta)} - 1] \end{aligned}$$

Remark 6.3. It is also possible to consider interacting particle systems where a bounded operator is added to A , without changing the core D . For example, one can consider

$$A_\theta f(\eta) = Af(\eta) + \sum_i c_i(\eta) [f(\theta_i \eta) - f(\eta)]$$

where θ_i is a shift: $(\theta_i \eta)_j = \eta_{i+j}$, and $\sum_i \|c_i\| < \infty$.

This includes processes like the environment process seen from a random walker in a dynamic random environment and the tagged particle process.

Our first goal is to equip D with a topology τ_D . The semi-norm $\|\cdot\|$ defined on D will be our starting point for τ_D as (6.3) implies $\|Af\| \leq \sup_x \sum_{\Lambda \ni x} c_\Lambda \|f\|$. Note that $\|\mathbb{1}\| = 0$, so $\|\cdot\|$ alone can not define a topology. We do have the following result.

Lemma 6.4. Let \mathcal{C} be the space of constant functions and let $\|\cdot\|_Q$ be the norm on the quotient space $C(E)/\mathcal{C}$. For $f \in D$, we have that $2\|f\|_Q \leq \|f\|$.

Proof. It is sufficient to prove the statement for local functions, because every $f \in D$ can be approximated by local f_n for which it holds that $\|f_n\| \rightarrow \|f\|$ and $\|f_n\|_Q \rightarrow \|f\|_Q$.

Suppose that f is a local function and let $\mathcal{D}(f) = \{x_1, \dots, x_n\}$. Now pick the function $f' \in D$ such that $f = f' + c$ for some $c \in \mathbb{R}$, such that the range of f'

is contained in $[0, 2\|f\|_Q]$. Pick η and ζ such that $f'(\eta) = 2\|f\|_Q$ and $f'(\zeta) = 0$. For $0 \leq k \leq n$ define $\Lambda_k = \{x_1, \dots, x_k\}$ and let ξ_k be equal to ζ on Λ_k and equal to η off Λ_k . Then it holds that

$$2\|f\|_Q = f'(\eta) = f'(\xi_0) = \sum_{k=0}^{n-1} f(\xi_k) - f(\xi_{k+1}) \leq \sum_{k=1}^n \Delta_f(x_k) = \|f\|.$$

□

The Lemma shows that one additional semi-norm is sufficient to topologise D . Let τ_D be the topology induced by $\|\cdot\|_D := \|\cdot\| + \|\cdot\|$.

Lemma 6.5. *(D, τ_D) is a separable Banach space.*

Proof. We start by proving that $(D, \|\cdot\|_D)$ is a Banach space, by using the following characterisation of completeness [4, Exercise III.4.2]. D is complete if and only if, for every sequence $f_n \in D$ such that

$$\sum_n \|f_n\|_D < \infty$$

the sum $\sum_{n=1}^N f_n$ converges in D .

So suppose that $\sum_n \|f_n\|_D < \infty$, then $\sum_n \|f\| < \infty$. Therefore, $\sum_n f_n \in C(E)$ as $(C(E), \|\cdot\|)$ is a Banach space. We need to show that $\sum_n f_n \in D$. By the definition of D , we need to check whether $\|\sum_n f_n\| < \infty$. But this follows from

$$\left\| \sum_n f_n \right\| \leq \sum_n \|f_n\| < \sum_n \|f_n\|_D < \infty.$$

So $(D, \|\cdot\|_D)$ is a Banach space and Banach spaces are always barrelled [27, Corollary 2 of Proposition 33.2].

We now prove separability of $(D, \|\cdot\|_D)$. For a finite box $\Lambda \subseteq S$, $w \in W^\Lambda$, and $\eta \in W^S$, define

$$\eta_\Lambda w_{\Lambda^c}(x) = \begin{cases} \eta_x & \text{for } x \in \Lambda \\ w_x & \text{for } x \in \Lambda^c \end{cases}.$$

Then define the local function $f_\Lambda \in D$ by $f_\Lambda(\eta) = f(\eta_\Lambda w_{\Lambda^c})$. As f is uniformly continuous, these local functions approximate f with respect to $\|\cdot\|_D$ as can be seen from the following computation.

$$\|f - f_\Lambda\|_D = \sum_{x \in \Lambda^c} \Delta_f(x) + \|f - f_\Lambda\| \rightarrow 0.$$

For a fixed and finite region $\Lambda \subseteq S$, the norm $\|\cdot\|_D$ restricted to the local functions depending on coordinates in Λ is equivalent to the sup norm. Therefore, this set of local functions is separable. By taking a sequence of finite regions $\Lambda_n \rightarrow S$. We obtain that the set of local functions is separable. By the argument above, every function in D can be approximated by local functions in the $\|\cdot\|$ semi-norm, so indeed $(D, \|\cdot\|_D)$ is separable. □

Proposition 6.6. *$(D, \|\cdot\|_D)$ satisfies Conditions 2.2 and 2.3.*

Proof. Conditions 2.3 (a) and (b) follow from Lemma 6.5. Conditions 2.2 and 2.3 (c) follows from a number of straightforward calculations using the seminorm $\|\cdot\|$.

By Theorem 6.2 (a) and (c), we obtain that $S(t) \in \mathcal{L}(D, \tau_D)$. An elementary calculation shows that for every $f \in D$ and $x \in S$, we have that $t \mapsto \Delta_{S(t)f}(x)$ is continuous. This implies, by using the Dominated convergence theorem and Theorem 6.2(c) that $t \mapsto S(t)f$ is continuous for $\|\cdot\|$. We conclude that $\{S(t)\}_{t \geq 0}$ is a strongly continuous semigroup for (D, τ_D) .

As $S(t)D \subseteq D$, Condition 2.2 (b') implies that also $V(t)D \subseteq D$. For a sequence of functions g_n that are uniformly bounded away from 0, we have that if $\|g_n - g\|_D \rightarrow 0$, then also $\|\log g_n - \log g\|_D \rightarrow 0$. Together with the continuity of $f \mapsto e^f$ by Condition 2.3 (c), we obtain the desired continuity properties of $V(t)$ from the properties of $S(t)$.

Condition 2.3 (e) is a direct consequence of Assumption (6.3) in Theorem 6.2. For (f), fix $f \in D$, then the function $\alpha \mapsto e^\alpha$ defined on $[-\|f\|, \|f\|]$ is Lipschitz continuous, with Lipschitz constant $e^{\|f\|}$. This means that $|e^\alpha - 1| \leq |\alpha|e^{\|f\|}$. Applying this to $\|Hf\|$, we obtain

$$\begin{aligned} \|Hf\| &\leq e^{2\|f\|} \|f\| \sum_{\Lambda} \left| \int c_{\Lambda}(\eta, d\zeta) [f(\eta^{\zeta}) - f(\eta)] \right| \\ &\leq e^{\|f\|} \|f\| \sup_x \sum_{\Lambda \ni x} c_{\Lambda} \end{aligned}$$

Using that for $x \geq 0$ $xe^x \leq e^{2x}$, (d) is satisfied by taking

$$\mathcal{N} := \left\{ f \in D \left| \|f\| \leq -\frac{1}{2} \log \left(\sup_x \sum_{\Lambda \ni x} c_{\Lambda} \right) \right. \right\}.$$

□

Proposition 6.6 implies that Theorem 2.8 holds for interacting particle systems where the derivative of the trajectory $t \mapsto \mu(t)$ lies in D' .

In the next section, we give a representation theorem for D' . Because we can always choose \mathcal{N} in Condition 2.3 such that it contains all constant functions, we can restrict our attention to $(D/\mathcal{C})'$, where \mathcal{C} is the space of constant functions. This is reasonable, because the only derivatives of a path of probability measures that we will find satisfy $\langle \mathbb{1}, u \rangle = 0$. We prove a representation theorem for $(D/\mathcal{C})'$.

6.2.1 A representation theorem for $((D/\mathcal{C})', \|\cdot\|)$

We identify the dual of D/\mathcal{C} the space of equivalence classes $D/\mathcal{C} \subseteq C(E)/\mathcal{C}$, where $\mathcal{C} := \{c\mathbb{1} \mid c \in \mathbb{R}\}$. Additionally, we equip D/\mathcal{C} with the norm $\|\cdot\|$, which is equivalent to the quotient norm $\|\cdot\|_{D, \mathcal{C}} = \|\cdot\|_{\mathcal{C}} + \|\cdot\|$ as $\|\cdot\| \leq \|\cdot\|_{D, \mathcal{C}} \leq \frac{3}{2} \|\cdot\|$ by Lemma 6.4.

We consider the dual of D/\mathcal{C} , which is equipped with the operator norm

$$\|\alpha\| = \sup_{f \in D/\mathcal{C}} \frac{|\langle f, \alpha \rangle|}{\|f\|}.$$

The goal of the following discussion is to identify both this dual space and its norm. First of all, the dual $(D/\mathcal{C})'$ can be seen as a subspace of functionals on D that are constant on the equivalence classes $f + \mathcal{C}$. Therefore,

$$\|\alpha\| = \sup_{f \in D} \frac{\langle f, \alpha \rangle}{\|f\|},$$

for α such that $\langle \alpha, \mathbb{1} \rangle = 0$.

We introduce some notation. For $\Lambda \subseteq S$, let $\mathcal{E}_\Lambda := \sigma(\eta_x \mid x \in \Lambda)$. Furthermore, $\tilde{\Pi}$ is the space of additive set functions α on the algebra $\mathcal{E}_a := \bigcup_{\Lambda: |\Lambda| < \infty} \mathcal{E}_\Lambda$, for which it holds that $\alpha(E) = 0$. Note that the σ -algebra \mathcal{E} is given by $\sigma(\mathcal{E}_a)$. For $\alpha \in \tilde{\Pi}$ and a finite subset $\Lambda \subseteq S$, we denote the restriction of α to \mathcal{E}_Λ by $P_\Lambda \alpha$ and we set $P_x := P_{\{x\}}$. Also, we define the function $\|\alpha\|_\Pi = \sup_x \|P_x \alpha\|_{TV}$ taking values in $[0, \infty]$.

Definition 6.7. Let Π be the set

$$\Pi := \left\{ \alpha \in \tilde{\Pi} \mid \|\alpha\|_\Pi < \infty \right\}.$$

It follows that Π is a vector space and that $\|\cdot\|_\Pi$ is a norm on Π . The following technical lemma enables us to show that $(\Pi, \|\cdot\|_\Pi)$ is a Banach space.

Lemma 6.8. For a finite set $T \subseteq S$: $\|P_\Lambda \alpha\|_{TV} \leq |\Lambda| \|\alpha\|_\Pi$.

Proof. Pick a local function f with dependence set $\mathcal{D}(f) = \{x_1, \dots, x_n\}$, $\sup_\eta f(\eta) = 2\|f\|_Q$ and $\inf_\eta f(\eta) = 0$. Pick ζ such that $f(\zeta) = 0$, and define for $k \leq n$ the sets $\Lambda_k = \{x_1, \dots, x_k\}$. For $\eta \in E$, let $\eta(k)$ be equal to ζ on Λ_k , and equal to η outside Λ_k . Furthermore, let $f_k(\eta) = f(\eta_k)$. Then it follows that

$$\begin{aligned} \int f d\alpha &= \int f_0(\eta) - f_n(\eta) d\alpha(\eta) \\ &= \sum_{k=0}^{n-1} \int f_k(\eta) - f_{k+1}(\eta) d\alpha(\eta) \\ &\leq \frac{1}{2} \sum_{k=0}^{n-1} \Delta_f(x_{k+1}) \|P_{x_{k+1}} \alpha\|_{TV} \\ &\leq \sum_{k=0}^{n-1} \|f\| \|P_{x_{k+1}} \alpha\|_{TV} \\ &\leq n \|f\| \|\alpha\|_\Pi. \end{aligned} \tag{6.5}$$

The bound obtained in line three of (6.5) is stronger than necessary, for this lemma, but we will use it again for the proof of Theorem 6.10. \square

Lemma 6.9. $(\Pi, \|\cdot\|_\Pi)$ is a Banach space.

Proof. We apply exercise III.4.2 in Conway [4] that states that $(\Pi, \|\cdot\|_\Pi)$ is complete if we can show for an arbitrary sequence $(\alpha_n)_{n \in \mathbb{N}}$ in Π such that $\sum_n \|\alpha_n\|_\Pi < \infty$, that the partial sums $\sum_n \alpha_n$ converge in Π . So pick a sequence α_n in Π such that $\sum_n \|\alpha_n\|_\Pi < \infty$. Furthermore, take a sequence of finite sets Λ_k that is increasing to S . By Lemma 6.8, we see that

$$\sum_n \|P_{\Lambda_k} \alpha_n\|_{TV} \leq |\Lambda_k| \sum_n \|\alpha_n\|_\Pi < \infty.$$

The space of measures on \mathcal{E}_{Λ_k} of bounded variation is a Banach space. Hence, $\alpha_{\Lambda_k} := \sum_n P_{\Lambda_k} \alpha_n$ exists and is a measure of bounded variation on \mathcal{E}_{Λ_k} . Furthermore, it is easy to see that this leads to a consistent sequence in k , so there exists a additive set function α on $\bigcup_{\Lambda: |\Lambda| < \infty} \mathcal{E}_{\Lambda}$, which, if restricted to finite regions, is a measure of bounded variation.

It follows that $\left\| \alpha - \sum_{k=1}^N \alpha_k \right\|_{\Pi} \rightarrow 0$, because

$$\left\| P_x \left(\alpha - \sum_{n=1}^N \alpha_n \right) \right\|_{TV} = \left\| P_x \left(\sum_{n=N+1}^{\infty} \alpha_n \right) \right\|_{TV} \leq \sum_{n=N+1}^{\infty} \|\alpha_n\|_{\Pi} \rightarrow 0.$$

□

We are now able to prove a representation theorem for $((D/\mathcal{C})', \|\cdot\|)$.

Theorem 6.10. $((D/\mathcal{C})', \|\cdot\|) = (\Pi, \frac{1}{2} \|\cdot\|_{\Pi})$, hence, $\|\alpha\| = \frac{1}{2} \sup_x \|P_x \alpha\|_{TV}$.

Proof. First, we show that $(D/\mathcal{C})'$ can be seen as a space of set functions. Take a finite set $\Lambda_0 \subseteq S$, then restricted the space $D_{\Lambda_0} := \{f \in D \mid \mathcal{D}(f) \subseteq \Lambda_0\}$ α is a continuous and linear function.

The space D_{Λ_0} with the topology induced by $\|\cdot\|$ is isomorphic to $C(W^{\Lambda_0})$ with the topology induced by $\|\cdot\|_Q$, as

$$2\|\cdot\|_Q \leq \|\cdot\| \leq 2|\Lambda_0| \|\cdot\|_Q.$$

Therefore, by the Riesz representation theorem, Theorem 7.10.4 in [1], it follows that for $f \in D_{\Lambda_0}$, $\alpha(f) = \langle f, \hat{\alpha}_{\Lambda_0} \rangle$ where $\hat{\alpha}_{\Lambda_0}$ is a measure of bounded variation on \mathcal{E}_{Λ_0} such that $\hat{\alpha}_{\Lambda_0}(E) = 0$. This can be done consistently for every finite set $\Lambda \subseteq S$, which implies that $\hat{\alpha}$ can be seen as a set function on $\bigcup_{\Lambda: |\Lambda| < \infty} \mathcal{E}_{\Lambda}$ for which the restriction to finite regions is a measure of bounded variation.

We proceed by showing that $\|\alpha\| \geq \frac{1}{2} \sup_x \|P_x \alpha\|_{TV}$. For $x \in S$, let $C(W^{\{x\}})$ be the set of continuous functions on W , but seen as local functions in D which depend only on the coordinate η_x .

$$\begin{aligned} \|\alpha\| &= \sup_{f \in D} \frac{|\langle f, \alpha \rangle|}{\|f\|} \geq \sup_{f \in C(W^{\{x\}})} \frac{|\langle f, \alpha \rangle|}{\|f\|} \\ &= \sup_{f \in C(W^{\{x\}})} \frac{|\langle f, \alpha \rangle|}{2\|f\|_Q} = \sup_{f \in C(W^{\{x\}})} \frac{|\langle f, \hat{\alpha} \rangle|}{2\|f\|_Q} = \frac{1}{2} \|P_x \hat{\alpha}\|_{TV} \end{aligned}$$

This means that the function $\Phi : (B/\mathcal{C})' \rightarrow \Pi$, mapping α to $\hat{\alpha}$, is well defined, injective and continuous. So, we identify α and $\hat{\alpha}$.

For the other inequality note that by continuity we can restrict the supremum to local functions:

$$\|\alpha\| = \sup_{f \text{ local}} \frac{|\langle f, \alpha \rangle|}{\|f\|}.$$

For local functions f , the result in Equation (6.5) yields:

$$\frac{|\langle f, \alpha \rangle|}{\|f\|} \leq \frac{\|f\| \frac{1}{2} \sup_x \|P_x \alpha\|_{TV}}{\|f\|} = \frac{1}{2} \sup_x \|P_x \alpha\|_{TV}.$$

This means that Φ is an isometry with respect to $\|\cdot\|$ and $\frac{1}{2} \|\cdot\|_{\Pi}$. We show that it is also surjective. Pick a local function f , then clearly $\alpha(f)$ is well

defined, because α restricted to $\mathcal{E}_{\mathcal{D}(f)}$ is a measure of bounded variation. By the calculation above we see that $|\langle f, \alpha \rangle| \leq \frac{1}{2} \|f\| \sup_x \|P_x \alpha\|_{TV}$. Hence, α defines a bounded linear functional on the local functions. Thus, it extends by continuity to a continuous linear functional on D/\mathcal{C} . \square

6.3 Diffusion processes on \mathbb{R}^d

We now show that our result partly reproduces the Dawson and Gärtner theorem [7]. First of all, we prove the result for a time-homogeneous case, but more importantly, we need to assume more regularity on the diffusion and drift terms. Let $C_0^m(\mathbb{R}^d)$ be the space of m times continuously differentiable functions, for which all derivatives up to order m are in $C_0(\mathbb{R}^d)$.

For every $x \in \mathbb{R}^d$, let $\{\sigma_{i,j}(x)\}_{i,j}$ be non-negative definite matrices, $\sigma_{i,j}(x)$ continuous in x . Denote with $a_{i,j}(x) = \sigma_{i,j}(x)\sigma_{i,j}(x)^T$. For each i , let $b_i \in C(\mathbb{R}^d)$. Define for every $f \in C_c^\infty(\mathbb{R}^d)$ the infinitesimal operator

$$Af(x) = \frac{1}{2} \sum_{i,j} a_{i,j}(x) \partial_i \partial_j f(x) + \sum_i b_i(x) \partial_i f(x).$$

Denote with

$$\begin{aligned} C_b^m(\mathbb{R}^d) &:= \{f \in C_b(\mathbb{R}^d) \mid D^\alpha f \in C_b(\mathbb{R}^d) \text{ if } |\alpha| \leq m\}, \\ C_0^m(\mathbb{R}^d) &:= \{f \in C_0(\mathbb{R}^d) \mid D^\alpha f \in C_0(\mathbb{R}^d) \text{ if } |\alpha| \leq m\}, \end{aligned}$$

and equip $C_0^m(\mathbb{R}^d)$ with the norm

$$\|f\|_m := \sum_{0 \leq |\alpha| \leq m} \|D^\alpha f\|.$$

Using the methods obtained to prove Theorem 8.2.5 in [12], we obtain the following theorem.

Theorem 6.11. *If $\sigma_{i,j}, b_i \in C_b^3(\mathbb{R}^d)$, then the closure of A generates a strongly continuous contraction semigroup on $C_0(\mathbb{R}^d)$. Additionally, $S(t)C_0^1(\mathbb{R}^d) \subseteq C_0^2(\mathbb{R}^d)$ and the restriction of $S(t)$ to $C_0^2(\mathbb{R}^d)$ is strongly continuous for $\|\cdot\|_2$.*

We calculate $A^g f$ and Hf for $f, g \in \mathcal{D} = C_0^2(\mathbb{R}^d)$. Again, the calculation of A^g gives us a new generator with a changed drift.

$$\begin{aligned} A^g f(x) &= Af(x) + \sum_{i,j} \frac{a_{i,j}(x) + a_{j,i}(x)}{2} \partial_j g(x) \partial_i f(x) \\ &= \frac{1}{2} \sum_{i,j} a_{i,j}(x) \partial_i \partial_j f(x) + \sum_i \left(b_i(x) + \sum_j \frac{a_{i,j}(x) + a_{j,i}(x)}{2} \partial_j g(x) \right) \partial_i f(x). \end{aligned}$$

Hf introduces a quadratic term:

$$Hf(x) = Af(x) + \frac{1}{2} \sum_{i,j} a_{i,j}(x) \partial_i f(x) \partial_j f(x). \quad (6.6)$$

As a corollary to Theorem 6.11, we obtain the next result.

Corollary 6.12. $(C_0^2(\mathbb{R}^d), \|\cdot\|_2)$ satisfies Condition 2.3.

Proof. Conditions (a) to (e) are straightforward to check. For (f), we put

$$\mathcal{N} := \left\{ f \in C_0^2(\mathbb{R}^d) \mid \begin{aligned} &\text{for all } x \in \mathbb{R}^d, \text{ we have } d \sup_i |b_i(x)| |\partial_i f(x)| \\ &+ \frac{d^2}{2} \sup_{i,j} |a_{i,j}(x)| (|\partial_i \partial_j f(x)| + |\partial_i f(x)| |\partial_j f(x)|) \leq 1 \end{aligned} \right\}.$$

Clearly, \mathcal{N} is closed, convex and balanced. We prove that \mathcal{N} is absorbing, which follows by showing that \mathcal{N} contains a ball $\{f \in C_0^2(\mathbb{R}^d) \mid \|f\|_2 \leq c\}$ for some c . Let $\bar{a} = \sup_{i,j} \sup_{x \in \mathbb{R}^d} |a_{i,j}(x)|$ and $\bar{b} = \sup_i \sup_{x \in \mathbb{R}^d} |b_i(x)|$. Pick $c > 0$ such that

$$\frac{d^2}{2} \bar{a}(c^2 + c) + d\bar{b}c \leq 1.$$

This choice implies that

$$\{f \in C_0^2(\mathbb{R}^d) \mid \|f\|_2 \leq c\} \subseteq \mathcal{N} \cap C^2(K_n).$$

We obtain that \mathcal{N} is a barrel and by construction of \mathcal{N} and the form of H , see (6.6), that $\sup_{f \in \mathcal{N}} \|Hf\| \leq 1$. Also, for $c \geq 1$, we obtain $\sup_{f \in c\mathcal{N}} \|Hf\| \leq c^2$. \square

A similar approach would give the result for $D = \mathcal{S}$ the space of rapidly decreasing smooth functions with its Fréchet space topology. This would need the extension for separable barrelled spaces in Condition 2.3. Note that \mathcal{S} is separable by the discussion following Proposition A.9 in [27] and barrelled by Corollary 1 of Proposition 33.2 in [27]

6.4 The Dawson and Gärtner theorem

As a consequence of the discussion above, we re-obtain a time-homogeneous and smooth version of Theorem 4.5 by Dawson and Gärtner [7].

Let (x_1, \dots, x_d) be the Euclidean coordinates. For $f \in C_0^2(\mathbb{R}^d)$, define $(\nabla f)^i = \sum_{j=1}^d a_{i,j}(\cdot) \frac{\partial f}{\partial x_j}$. Then it follows from Equation (6.6) that $\langle Hf, \mu \rangle = \langle Af, \mu \rangle + \frac{1}{2} \langle |\nabla f|^2, \mu \rangle$.

We introduce two new spaces,

$$\begin{aligned} D_\mu &:= \{f \in D = C_0^2(\mathbb{R}^d) \mid \langle |\nabla f|^2, \mu \rangle \neq 0\} \\ T_\mu &:= \left\{ \alpha \in C_0^2(\mathbb{R}^d)' \mid \|\alpha\|_\mu < \infty \right\}, \end{aligned}$$

where $\|\cdot\|_\mu$ is defined on $C_0^2(\mathbb{R}^d)'$ by

$$\|\alpha\|_\mu := \sup_{f \in D_\mu} \frac{|\langle f, \alpha \rangle|^2}{\langle |\nabla f|^2, \mu \rangle}.$$

The next proposition shows the connection between Theorem 2.8 and Theorem 4.5 by Dawson and Gärtner [7].

Proposition 6.13. *If $\mathcal{L}(\mu, \alpha) < \infty$, then $\alpha \in T_\mu$ and $\mathcal{L}(\mu, \alpha) = \frac{1}{2} \|\alpha - A'\mu\|_\mu$. As a consequence for a trajectory $\nu \in \mathcal{AC}$*

$$\int_0^\infty \mathcal{L}(\nu(s), \dot{\nu}(s)) ds = \frac{1}{2} \int_0^\infty \|\dot{\nu}(s) - A'\nu(s)\|_{\nu(s)} ds.$$

Proof. Pick $\mu \in \mathcal{P}(\mathbb{R}^d)$ and $\alpha \in C_0^2(\mathbb{R}^d)'$ such that $\mathcal{L}(\mu, \alpha) < \infty$. Define $\hat{\alpha} = \alpha - A'\mu$. Consider

$$\begin{aligned} \mathcal{L}(\mu, \alpha) &= \sup_{f \in C_0^2(\mathbb{R}^d)} \{ \langle f, \alpha \rangle - \langle Hf, \mu \rangle \} \\ &= \sup_{f \in C_0^2(\mathbb{R}^d)} \left\{ \langle f, \alpha \rangle - \langle Af, \mu \rangle - \frac{1}{2} \langle |\nabla f|^2, \mu \rangle \right\} \\ &= \sup_{f \in C_0^2(\mathbb{R}^d)} \left\{ \langle f, \hat{\alpha} \rangle - \frac{1}{2} \langle |\nabla f|^2, \mu \rangle \right\} \\ &= \sup_{f \in C_0^2(\mathbb{R}^d)} \sup_{c \in \mathbb{R}} \left\{ c \langle f, \hat{\alpha} \rangle - c^2 \frac{1}{2} \langle |\nabla f|^2, \mu \rangle \right\} \end{aligned} \quad (6.7)$$

By assumption, the supremum in the equation above is finite. Then if $f \in D_\mu^c$, it must be that $\langle f, \hat{\alpha} \rangle = 0$. Therefore, these f yield 0 as an argument in the supremum.

For a given $f \in D_\mu$, optimising over c yields $c = \frac{\langle f, \hat{\alpha} \rangle}{\langle |\nabla f|^2, \mu \rangle}$. Therefore, we can rewrite Equation (6.7) as

$$\mathcal{L}(\mu, \alpha) = 0 \vee \frac{1}{2} \sup_{f \in D_\mu} \frac{|\langle f, \hat{\alpha} \rangle|^2}{\langle |\nabla f|^2, \mu \rangle} = \frac{1}{2} \|\alpha - A'\mu\|_\mu.$$

□

7 Appendix: Decomposition of the rate function on product spaces

In this appendix, (E, d) is a complete separable metric space.

Suppose \mathbb{P} is the law of a Markov process on $D_E(\mathbb{R}^+)$. Suppose that the sequence $(L_n^{X(0)}, \dots, L_n^{X(t_k)})$ satisfies the large deviation principle on $\mathcal{P}(E)^{k+1}$. The following lemma is a multidimensional version of exercise 6.2.26 of Dembo and Zeitouni [8].

Lemma 7.1. *The large deviation rate function $I[0, t_1, \dots, t_k]$ of the LDP of the sequences $(L_n^{X(0)}, \dots, L_n^{X(t_k)})$ on $\mathcal{P}(E)^{k+1}$ is given by*

$$\begin{aligned} I[0, t_1, \dots, t_k](\nu_0, \dots, \nu_k) \\ = \sup_{f_0, \dots, f_k \in C_b(E)} \sum_{i=0}^k \langle f_i, \nu_i \rangle - \log \mathbb{E} \left[e^{f_0(X(0)) + \dots + f_k(X(t_k))} \right]. \end{aligned} \quad (7.1)$$

Also, we can restrict to a smaller class of functions, see [8, Definition 4.4.7 and exercise 4.4.14].

Corollary 7.2. *The supremum over $C_b(E)$ in (7.1) can be restricted to any class of functions M that separates points and is closed under taking point-wise minima. In particular, this holds for $C_0(E)$ if E is locally compact.*

Denote with $V(s, t)f(x) = \log \mathbb{E}_{X(s)=x} [e^{f(X(t))}]$ and put

$$I_{t_1, t_2}(\nu_1 | \nu_0) = \sup_{f \in M} \langle f, \nu_1 \rangle - \langle V(t_1, t_2)f, \nu_0 \rangle.$$

Clearly, if X is a time-homogeneous process, we can simplify to $V(t-s) := V(s, t)$ and $I_{t_2-t_1} := I_{t_1, t_2}$.

Proposition 7.3. *Let $M \subseteq C_b(E)$ be a set of functions that separates points, and which is closed under taking point-wise minima. Denote with $V(s, t)f(x) = \log \mathbb{E}_{X(s)=x} [e^{f(X(t))}]$ and let M be such that for every $t \geq 0$: $V(t)M \subseteq M$. Define*

$$I_{t_1, t_2}(\nu_1 | \nu_0) = \sup_{f \in M} \langle f, \nu_1 \rangle - \langle V(t_1, t_2)f, \nu_0 \rangle.$$

Then, it holds that

$$I[0, t_1, \dots, t_k](\nu_0, \dots, \nu_k) = I_0(\nu_0) + \sum_{i=1}^k I_{t_{i-1}, t_i}(\nu_i | \nu_{i-1}).$$

Proof. The proof consists of two parts. First, we prove that

$$I[0, t_1](\nu_0, \nu_1) = I_0(\nu_0) + I_{0, t_1}(\nu_1 | \nu_0).$$

In line 2, we have taken a conditional expectation to reduce $f_1(X(t_1))$ to $V(0, t_1)f_1(X(0))$.

$$\begin{aligned} I[0, t_1](\nu_0, \nu_1) &= \sup_{f_0, f_1 \in M} \langle f_0, \nu_0 \rangle + \langle f_1, \nu_1 \rangle - \log \mathbb{E} \left[e^{f_0(X(0)) + f_1(X(t_1))} \right] \\ &= \sup_{f_0, f_1 \in M} \langle f_0, \nu_0 \rangle + \langle f_1, \nu_1 \rangle - \log \mathbb{E} \left[e^{f_0(X(0)) + V(0, t_1)f_1(X(0))} \right] \\ &= \sup_{f_1 \in M} \left\{ \langle f_1, \nu_1 \rangle - \langle V(0, t_1)f_1, \nu_0 \rangle \right. \\ &\quad \left. + \sup_{f_0 \in M} \left\{ \langle f_0 + V(0, t_1)f_1, \nu_0 \rangle - \log \mathbb{E} \left[e^{f_0(X(0)) + V(0, t_1)f_1(X(0))} \right] \right\} \right\} \\ &= \sup_{f_1 \in M} \langle f_1, \nu_1 \rangle - \langle V(0, t_1)f_1, \nu_0 \rangle + I_0(\nu_0) \\ &= I_0(\nu_0) + I_{0, t_1}(\nu_1 | \nu_0) \end{aligned}$$

In line 5, we use that $f_0 + V(0, t_1)f_1 \in M$.

The proposition follows by induction from the statement

$$I[0, t_1, \dots, t_k](\nu_0, \dots, \nu_k) = I[0, t_1, \dots, t_{k-1}](\nu_0, \dots, \nu_{k-1}) + I_{t_{k-1}, t_k}(\nu_k | \nu_{k-1}),$$

which follows by the same argument as above. \square

8 Appendix: Souslin spaces, barrelled spaces, and Gelfand integration

8.1 Barrelled spaces and Gelfand integration

Definition 8.1. A locally convex space \mathcal{X} is called barrelled if every barrel is a neighbourhood of 0. A set S is a barrel if it is convex, balanced, absorbing

and closed. S is balanced if we have the following: if $x \in S$ and $\alpha \in \mathbb{R}$, $|\alpha| \leq 1$ then $\alpha x \in S$. S is absorbing if for every $x \in \mathcal{X}$ there exists a $r \geq 0$ such that if $|\alpha| \geq r$ then $x \in \alpha S$.

For example, Banach, Fréchet and LF(limit Fréchet) spaces are barrelled [27, Chapter 33]. The space of Schwartz functions is Fréchet and the space $C_c^\infty(\mathbb{R}^d)$ with its usual topology is LF.

The importance of barrelled spaces follows from the fact that the closed graph theorem holds for them [24, Proposition 7.1.11], [26, Theorem VI.7].

Theorem 8.2 (Closed graph theorem). *Let \mathcal{X} be a barrelled locally convex space, and let F be a Fréchet space. Suppose that $T : F \rightarrow \mathcal{X}$ is a linear operator with closed graph in $F \times \mathcal{X}$, then T is continuous.*

The closed graph theorem is of importance for integration of functions with values in the dual of a barrelled space. Let $(\Omega, \mathcal{F}, \mu)$ be a complete and finite measure space, and let \mathcal{X} be a barrelled space with continuous dual \mathcal{X}' . We equip \mathcal{X}' with $\sigma(\mathcal{X}', \mathcal{X})$, the weak* topology.

Definition 8.3. A function $f : \Omega \rightarrow \mathcal{X}'$ is called weak* measurable if the scalar function

$$\omega \mapsto \langle x, f(\omega) \rangle$$

is \mathcal{F} measurable for every $x \in \mathcal{X}$. Such a function f is called *Gelfand* or weak* integrable if $\langle x, f \rangle \in \mathcal{L}^1(\Omega, \mathcal{F}, \mu)$ for every $x \in \mathcal{X}$.

For Gelfand integrable functions, we obtain, using the Closed graph theorem, the following result [10, pages 52-53].

Theorem 8.4. *Let \mathcal{X} be a barrelled space and $(\Omega, \mathcal{F}, \mu)$ a complete and finite measure space. For every measurable set $A \in \mathcal{F}$ and Gelfand integrable function $f : \Omega \rightarrow \mathcal{X}'$, there exists a unique $x'_A \in \mathcal{X}'$ such that*

$$\langle x, x'_A \rangle = \int_A \langle x, f(\omega) \rangle \mu(d\omega)$$

for all $x \in \mathcal{X}$. This element x'_A will be denoted by $\int_A f d\mu$.

8.2 Souslin spaces

Definition 8.5. A space (Y, τ_Y) is called Souslin, if $Y = f(X)$ for some complete separable metric space (X, τ_X) and some continuous function $f : (X, \tau_X) \rightarrow (Y, \tau_Y)$.

For more background on Souslin spaces, see Chapters 6 and 7 in Bogachev [1].

Lemma 8.6. *Let (X, τ) be a separable barrelled locally convex Hausdorff space and T a barrel in (X, τ) . Then $(\bigcup_n nT^\circ, wk^*) \subseteq (X', wk^*)$ is a Souslin space.*

In particular, as the unit ball in a Banach space B is a barrel, the dual (B', wk^*) of separable Banach space is Souslin.

Proof. As (X, τ) is barrelled, T is a neighbourhood of 0. Consequentially, T° is an equi-continuous set in (X^*, wk^*) by 21.3.(1) in Köthe [19]. By the Bourbaki-Alaoglu theorem, 20.9.(4) [19], this set is weak* compact. Furthermore, by 39.4.(7) in [20], T° is metrisable. (T°, wk^*) is compact and metric, which implies that it is complete separable metric and as a consequence Souslin. We can do the same for $n\mathcal{N}^\circ$ for every $n \in \mathbb{N}$, so we obtain that $(\bigcup_n n\mathcal{N}^\circ, wk^*)$ is Souslin [1, Theorem 6.6.6]. \square

Acknowledgement The author thanks Frank Redig and Jan van Neerven for reading the manuscript and numerous valuable comments. The author also thanks an anonymous referee for pointing out a mistake and suggestions that improved the text. The author is supported by The Netherlands Organisation for Scientific Research (NWO), grant number 600.065.130.12N109.

References

- [1] Vladimir I. Bogachev. *Measure Theory*. Springer-Verlag, 2007.
- [2] Amarjit Budhiraja, Paul Dupuis, and Markus Fischer. Large deviation properties of weakly interacting processes via weak convergence methods. *Ann. Probab.*, 40(1):74–102, 2012.
- [3] K. Netočný C. Maes and B. Wynants. On and beyond entropy production; the case of Markov jump processes. *Markov Processes and Related Fields*, 14:445–464, 2008.
- [4] John B. Conway. *A Course in Functional Analysis*. Springer, second edition, 2007.
- [5] M. G. Crandall and T. M. Liggett. Generation of semi-groups of nonlinear transformations on general banach spaces. *American Journal of Mathematics*, 93(2):pp. 265–298, 1971.
- [6] I. Csiszár. I -divergence geometry of probability distributions and minimization problems. *Ann. Probability*, 3:146–158, 1975.
- [7] Donald A. Dawson and Jürgen Gärtner. Large deviations from the McKean-Vlasov limit for weakly interacting diffusions. *Stochastics*, 20(4):247–308, 1987.
- [8] Amir Dembo and Ofer Zeitouni. *Large Deviations Techniques and Applications*. Springer-Verlag, second edition, 1998.
- [9] Jean-Dominique Deuschel and Daniel W. Stroock. *Large deviations*, volume 137 of *Pure and Applied Mathematics*. Academic Press, Inc., Boston, MA, 1989.
- [10] J. Diestel and J. J. Uhl, Jr. *Vector measures*. American Mathematical Society, 1977.
- [11] J. L. Doob. *Classical potential theory and its probabilistic counterpart*, volume 262 of *Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences]*. Springer-Verlag, New York, 1984.

- [12] Stewart N. Ethier and Thomas G. Kurtz. *Markov processes: Characterization and Convergence*. Wiley, 1986.
- [13] Jin Feng and Thomas G. Kurtz. *Large Deviations for Stochastic Processes*. American Mathematical Society, 2006.
- [14] Shui Feng. Large deviations for empirical process of mean-field interacting particle system with unbounded jumps. *Annals of Probability*, 22(4):2122–2151, 1994.
- [15] Shui Feng. Large deviations for Markov processes with mean field interaction and unbounded jumps. *Probability Theory and Related Fields*, 100:227–252, 1994.
- [16] Hans Föllmer. Random fields and diffusion processes. In *École d’Été de Probabilités de Saint-Flour XV–XVII, 1985–87*, volume 1362 of *Lecture Notes in Math.*, pages 101–203. Springer, Berlin, 1988.
- [17] Hans Föllmer and Nina Gantert. Entropy minimization and Schrödinger processes in infinite dimensions. *Ann. Probab.*, 25(2):901–926, 04 1997.
- [18] Benton Jamison. The Markov processes of Schrödinger. *Zeitschrift für Wahrscheinlichkeitstheorie und Verwandte Gebiete*, 32(4):323–331, 1975.
- [19] Gottfried Köthe. *Topological Vector Spaces I*. Springer-Verlag, 1969.
- [20] Gottfried Köthe. *Topological Vector Spaces II*. Springer-Verlag, 1979.
- [21] Christian Léonard. Large deviations for long range interacting particle systems with jumps. *Annales de l’institut Henri Poincaré (B) Probabilités et Statistiques*, 31(2):289–323, 1995.
- [22] Christian Léonard. A survey of the Schrödinger problem and some of its connections with optimal transport. *Discrete Contin. Dyn. Syst.*, 34(4):1533–1574, 2014.
- [23] Thomas M. Liggett. *Interacting Particle Systems*. Springer-Verlag, 1985.
- [24] J. Bonet P. Pérez Carreras. *Barrelled Locally Convex Spaces*. North-Holland Publishing Co., Amsterdam, 1987.
- [25] Zbigniew Palmowski and Tomasz Rolski. A technique for exponential change of measure for Markov processes. *Bernoulli*, 8(6):767–785, 2002.
- [26] A.P. Robertson and W. Robertson. *Topological Vector Spaces*. Cambridge University Press, second edition, 1973.
- [27] François Trèves. *Topological Vector Spaces, Distributions and Kernels*. Academic Press, New York-London, 1967.